



### **III. THE PHASE-SHIFTING POINT DIFFRACTION INTERFEROMETER**



# 4

## *The Phase-Shifting Point Diffraction Interferometer*

---

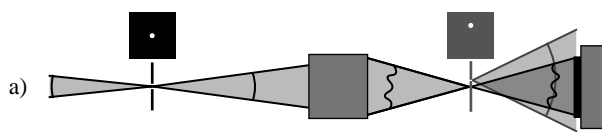
<b>4.1 INTRODUCTION AND MOTIVATION</b>	<b>48</b>
<b>4.2 PS/PDI DESCRIPTION</b>	<b>49</b>
<b>4.3 CONFIGURATIONS OF THE PS/PDI</b>	<b>49</b>
<b>4.4 COMPETING INTERFEROMETER DESIGNS</b>	<b>51</b>
<b>4.5 INTENSITY AND EFFICIENCY CONSIDERATIONS</b>	<b>52</b>
<b>4.6 CHOOSING THE OPTIMAL PINHOLE SIZE</b>	<b>56</b>
4.6.1 Reference Wave Uniformity	
4.6.2 A Simple Approach to Pinhole Spatial Filtering Considerations	

#### 4.1 INTRODUCTION AND MOTIVATION

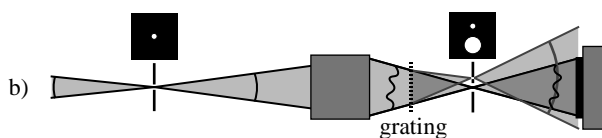
Experience with the EUV PDI yielded an understanding of the serious limitations of the conventional point-diffraction interferometer for high-accuracy measurements, and led ultimately to the development of a novel point diffraction interferometer design capable of greatly improved throughput and possessing the capacity for phase-shifting interferometry — the *Phase-Shifting Point Diffraction Interferometer* (PS/PDI) (Medeck et al. 1996, Goldberg et al. 1997, Tejnli et al. 1997) This chapter describes the design and basic operational principles of the PS/PDI. The advantages of the PS/PDI are described relative to other common-path EUV interferometer designs.

High-accuracy wavefront measurement with the Conventional PDI (hereafter referred to as simply PDI) is hindered by several factors. In the PDI design (Fig. 1(a)), the reference wavefront is generated by diffraction from a sub-resolution *reference pinhole* in a partially-transmitting membrane. The test beam is formed from the light that is transmitted through the membrane, containing the aberrations of the optic under test. Since the reference beam is sampled from the test beam, there is no available means to introduce a controllable relative phase-shift between the two; therefore static fringe pattern analysis methods must be used. The reliability of such analyses is limited in the presence of mid- and high-spatial frequency variations of the test beam intensity. Furthermore, significant lateral displacements of the reference pinhole from focus, typically 10-25 times  $\lambda/\text{NA}$  (1-2  $\mu\text{m}$  in EUV interferometry), are required to generate enough fringes for static fringe pattern analysis. Such displacement greatly decreases the amount of pinhole-diffracted light available for the reference wavefront. Consequently, to match the intensities of the two waves, and to provide good fringe contrast, the membrane must *significantly* attenuate the test wavefront; this reduces the overall throughput, or efficiency, of the interferometer. Such necessary beam attenuation may make alignment and measurement difficult by pushing the required single-image exposure time into the range of several minutes.

Conventional PDI



One Example of the PS/PDI



**Figure 1.** Schematic representations of (a) the Conventional Point Diffraction Interferometer (PDI), and (b) the closely related Phase-Shifting Point Diffraction Interferometer (PS/PDI). Both systems require coherent illumination of the optic under test. The PDI uses a partially-transmitting membrane and a sub-resolution pinhole to sample the aberrated test beam and produce a reference wavefront. The PS/PDI utilizes a low-angle beamsplitter to divide the test beam into multiple separate beams in the image plane. One beam passes through a large open window in an opaque image-plane membrane. A second beam is focused onto a sub-resolution pinhole and produces a reference wavefront.

## 4.2 PS/PDI DESCRIPTION

In the PS/PDI designs, of which one example is shown in Fig. 1(b), a small-angle beamsplitter (such as a coarse grating) is employed to separate the test and reference beams, forming multiple foci in the image-plane. Using a two-pinhole spatial filter in the image-plane, two beams are selected: one beam passes through a large *window* in an opaque membrane, while another beam is diffracted by a sub-resolution *reference pinhole* placed at the center of the focal spot.

This design overcomes several of the limitations of the conventional PDI. Translating the grating beamsplitter perpendicular to the grating ruling introduces a controllable relative phase-shift between the test and reference wavefronts, facilitating *phase-shifting interferometry* (PSI), a powerful category of data analysis techniques. Additionally, the centered reference pinhole and the large open window lead to an overall throughput increase of at least two orders of magnitude compared to the conventional PDI.

There are many ways in which the PS/PDI may be used, and several available variations on the basic design. Besides the ever-present concern about the size of the reference pinhole, the grating beamsplitter may be placed in several available locations. The position and pitch of the grating determine the separation of the test and reference beams in the image-plane. The appropriate separation depends on the quality of the optical system under test and on the desired mid-spatial frequency resolution of the interferometer. There are also advantages and disadvantages related to the selection of which of the diffracted orders becomes the test and reference beams. These issues and others are addressed in the following sections.

## 4.3 CONFIGURATIONS OF THE PS/PDI

One central advantage of the PS/PDI over the PDI is that the reference pinhole is centered on the brightest part of the focused illumination, greatly enhancing the amount of transmitted light in the reference beam. In any configuration of the PS/PDI, one primary motivation is to deliver the highest available flux to the reference pinhole. Since the pinhole acts as a spatial filter, removing any aberration in the reference beam, the primary quality of concern for the reference beam is simply its focused intensity. In principle, the beamsplitter may be placed in any available position ahead of the image-plane.

Figure 2(a) shows the conventional PDI alongside several configurations of the PS/PDI with a grating beamsplitter and one using a glancing-incidence mirror. Figures 2(b) and (c) show two similar configurations with the grating placed either before or after the test optic. When the wavefront division occurs ahead of the test optic, the multiple beams will travel along different paths through the system; in extreme cases, consideration must be given to the fact that apertures in the system may block all or part of the beams.

In any of the PS/PDI configurations, a choice must be made as to which beam is the test beam and which beam is filtered to become the reference beam. These are called the *first-order reference* and the *zeroth-order reference* configurations, denoting which beam becomes the reference. Since the beam separa-

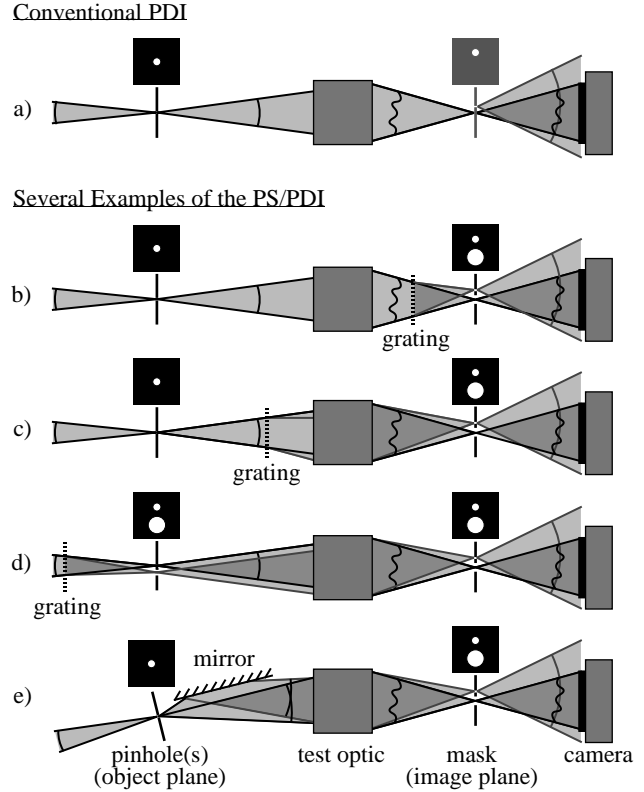
tion is typically small in the image-plane, switching between the two configurations is usually a trivial matter. However, intensity and beam quality issues, discussed here and in Section 5.10, often motivate the use of one configuration over the other.

Figure 2(d) shows the *upstream grating configuration* applicable in circumstances where the illuminating beam is of high quality. In this design, two similar two-pinhole spatial filters are used: one in the object plane, and a second in the image-plane. A grating placed ahead of the object pinholes separates the converging test and reference beams. A small object pinhole filters the test beam, guaranteeing a spatially coherent, spherical illumination wavefront. The reference beam, however, passes through a large window in the object plane and is filtered by the reference pinhole in the image plane. In bypassing the spatial filter pinhole in the object plane, the reference

beam reaches the image plane with much greater intensity than in the other PS/PDI configurations.

Although advantageous in this regard, the upstream grating configuration requires that the illuminating beam incident on the object plane be of sufficiently high quality (i.e. nearly diffraction-limited) to be well separable. This requirement precludes its implementation in EUV applications where the illuminating optics typically are not of near-diffraction-limited quality.

Depending on the operational beam wavelength, there may be several available ways of achieving the required wavefront division. The grating systems are convenient because a relative phase-shift between any two diffracted orders is induced by a simple lateral translation of one component, and is therefore straightforward to implement and control. Another system, shown in Fig. 2(e), is reminiscent of Lloyd's mirror (Born and Wolf 1980:262-263), where a glancing-incidence mirror is used to *fold* the illuminating beam onto itself over the entrance pupil. Here, the NA of the illuminating wavefront must be of more than double the object-side NA of the system. The implementation of phase-shifting in this configuration is problematic, if it is possible.



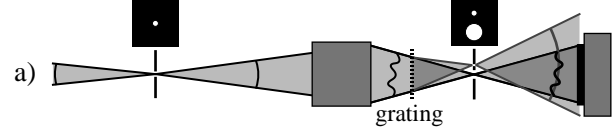
**Figure 2.** The PDI (a) is compared to four configurations of the PS/PDI (b-e). (b) and (c) are different only in the placement of the grating beamsplitter. In configuration (d) the beam is split upstream of the object pinhole, and similar window-plus-pinhole masks are placed in both the object- and image-planes. Configuration (e) shows a different mechanism of beam-splitting, using a Lloyd mirror.

#### 4.4 COMPETING INTERFEROMETER DESIGNS

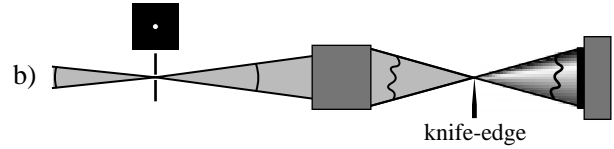
In addition to the two point-diffraction interferometer designs used in this research, two other common-path interferometer designs have been implemented for EUV optics testing: the *Knife-Edge* or *Foucault Test* (Foucault 1858, 1859), and the grating-based *Lateral Shearing Interferometer* (LSI), or *Ronchi Interferometer* (Ronchi 1923, 1964), shown in Fig. 3.

The simple-to-perform Knife-Edge test involves placing a high-quality opaque edge at the focus of the optical system under test. By blocking some of the aberrated rays, the resultant far-field intensity pattern reveals the slope of the wavefront. This test was successfully employed in the alignment of an EUV Schwarzschild objective (Ray-Chaudhuri 1994). For high-accuracy applications, the advantages of the Foucault test in simplicity, sensitivity, and high-efficiency are outweighed by the difficulty in performing accurate analysis of the data.

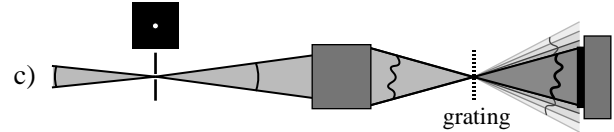
##### One Example of the PS/PDI



##### Knife-Edge (Foucault)



##### Grating LSI (Ronchi)



**Figure 3.** Schematic drawings of the PS/PDI and two common-path interferometers that have been used in EUV interferometry.

The grating-based LSI is another test convenient because of its relative simplicity. This interferometer design has also been used to test 10× Schwarzschild objectives identical in design to those under investigation with the PS/PDI (Ray-Chaudhuri 1997, Wood et al. 1997). A coarse grating is placed near the focus of the optic under test. The grating divides the beam into multiple, overlapping orders which are *sheared* angularly in the direction perpendicular to the grating rulings. In a typical *shearing interferometer*, the interference of *two* slightly-displaced overlapping beams reveals the wavefront slope along the direction of the shear. Here, analysis is complicated by the presence of multiple overlapping beams. The shear angle is determined by the grating pitch; the important parameter is the ratio of the shear angle to the NA. The amount of shear dictates the slope of the measured wavefront, and therefore largely determines the sensitivity of the technique. Using grating translation to induce phase-shifting into the measurements, wavefront slope data is gathered along two shear directions, and the two separate measurements must be reconciled to reconstruct the wavefront. Although the success of this technique has been demonstrated, its applicability to high accuracy wavefront measurement is still under investigation.

Both of these interferometer designs have advantages over the PS/PDI: higher efficiency, because a second spatial-filter pinhole is not used, and ease of alignment, because the placement of a tiny pinhole onto the beam focus is unnecessary. There are fewer critical components and stages, and those compo-



nents are easier to obtain. Acknowledging these advantages, however, the PS/PDI design has many positive attributes not possessed by the other two. In generating a single reference wavefront by pinhole diffraction, extremely high accuracies may be achieved. The interference data that is collected enables measurement of the wavefront itself, not the wavefront slope, so analysis is more straightforward and uncertainties are greatly reduced. Because of the high-brightness synchrotron source in use for the EUV PS/PDI experiments, the relatively lower efficiency of the PS/PDI has not presented any significant experimental disadvantage.

#### 4.5 INTENSITY AND EFFICIENCY CONSIDERATIONS

The relative efficiencies of the PDI and PS/PDI configurations vary widely. This issue may be of foremost concern in circumstances where the available intensity of coherent illumination is limited. The efficiency dictates how much time is required to conduct interferometric measurements. Here, to illustrate this variation, a few simplified assumptions about the loss mechanisms are applied to a side-by-side comparison of the different point diffraction interferometer configurations. The relative efficiencies of the PS/PDI, the Knife-Edge Test, and the LSI are also compared.

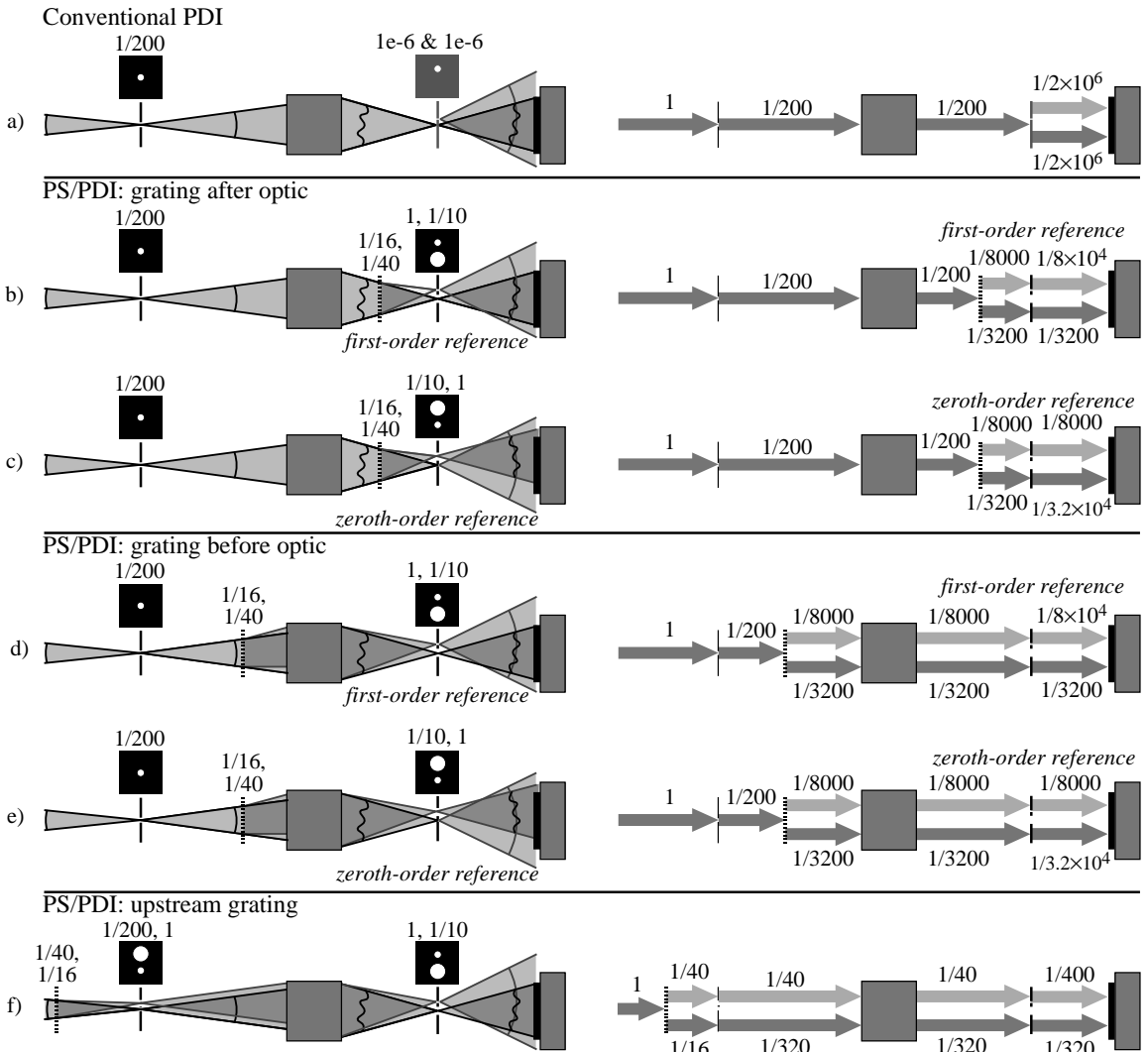
The EUV PS/PDI, configured for the testing of a 10× Schwarzschild objective as described in this thesis, will serve as a model for this exercise. Experimental characteristics of the synchrotron beamline source and several of the interferometer's components are applied here. The inherent efficiency of the test optic will affect each of these common-path interferometers in the same way and is therefore neglected in this discussion.

The object pinhole is illuminated by a beam of marginal quality, forming a focal spot of approximately  $50\text{-}\mu\text{m}^2$  area (at 0.008 NA). A  $0.5\text{-}\mu\text{m}$ -diameter object pinhole transmits approximately 1/200th of the incident illumination. Assume in this discussion that for high-quality optics, the image-plane reference pinholes transmit 1/10th of the incident illumination; also assume that the large *window* pinholes of the PS/PDI have 100% transmission. When aberrated optical systems are tested, the size of the focal spot increases and transmission through the reference pinhole is reduced. This does not affect the efficiencies of the Knife-Edge or LSI test, but it *significantly* affects the assumptions made here about transmission through the reference pinhole.

Assume for simplicity that the transmission gratings are binary: alternating opaque and transparent stripes of 1:1 line-to-space ratio. Phase-gratings and gratings with a line-to-space ratio other than 1:1 could be used to improve throughput or to match the relative intensities of the test and reference beams; but in this example, only the simple grating will be considered. For such an ideal grating, the intensity transmitted into the zeroth-order beam is 1/4, and the intensity in one of the the first-order beams is  $1/\pi^2$ .

(A good rule of thumb for such gratings is that the ratio of the intensities of the first order to the zeroth order is  $4/\pi^2 \approx 40\%$ .) The gratings used in EUV interferometry are typically supported by a 1000-Å silicon-nitride membrane, with a transmission of 1/4. Thus the total intensity transmitted into the zeroth- and first-orders is 1/16, and  $1/4\pi^2 \approx 1/40$  respectively.

Regarding the conventional PDI, assume for example that 20 fringes are desired, necessitating a lateral displacement of the reference pinhole by approximately  $10 \lambda/\text{NA}$ . If the area of this displaced pinhole is one-quarter of the central Airy disk area (a desirable size), then the amount of light transmitted through this pinhole in a high-quality optical system can be on the order of  $10^{-5}$  to  $10^{-6}$ . To balance the



**Figure 4.** An efficiency comparison of the different point diffraction interferometer designs. On the left, the approximate efficiency of each element is shown. The efficiency of the optic itself, the same in all configurations, is omitted. On the right are shown the approximate integrated intensities of the test and reference beams as they propagate through the interferometers. The first-order reference and the zeroth-order reference configurations are also compared: in (b) and (c), and in (d) and (e). This side-by-side comparison reveals the efficiency advantages of some configurations over others. Numbers are given in Table 1.

two illuminating beams, the transmission of the semi-transparent membrane must be of this same order of magnitude. In practice, it is possible that an optic with large mid-spatial-frequency errors may scatter more radiation away from the center and into the vicinity of the displaced pinhole. Furthermore, if the reference pinhole is larger than it should be, the flux transmitted into the reference beam may be closer to  $10^{-4}$  than to  $10^{-6}$ .

Figure 4 compares the efficiency of each PS/PDI design. On the left are schematic representations of interferometers. Above each of the essential components, the approximate efficiencies (photons *out* versus photons *in*) used in the calculations are shown. On the right are representations of the integrated test and reference beam intensities in each segment of the interferometers. There are several important values to consider. These require the following definition: the test and reference beams combine to form a *stationary intensity* and a *modulated intensity* that can be represented as

$$I = I_{stationary} + I_{modulated} \cos \Phi, \quad (1)$$

with  $\Phi$  as the arbitrary optical path difference in radians. This and the definition of fringe contrast are

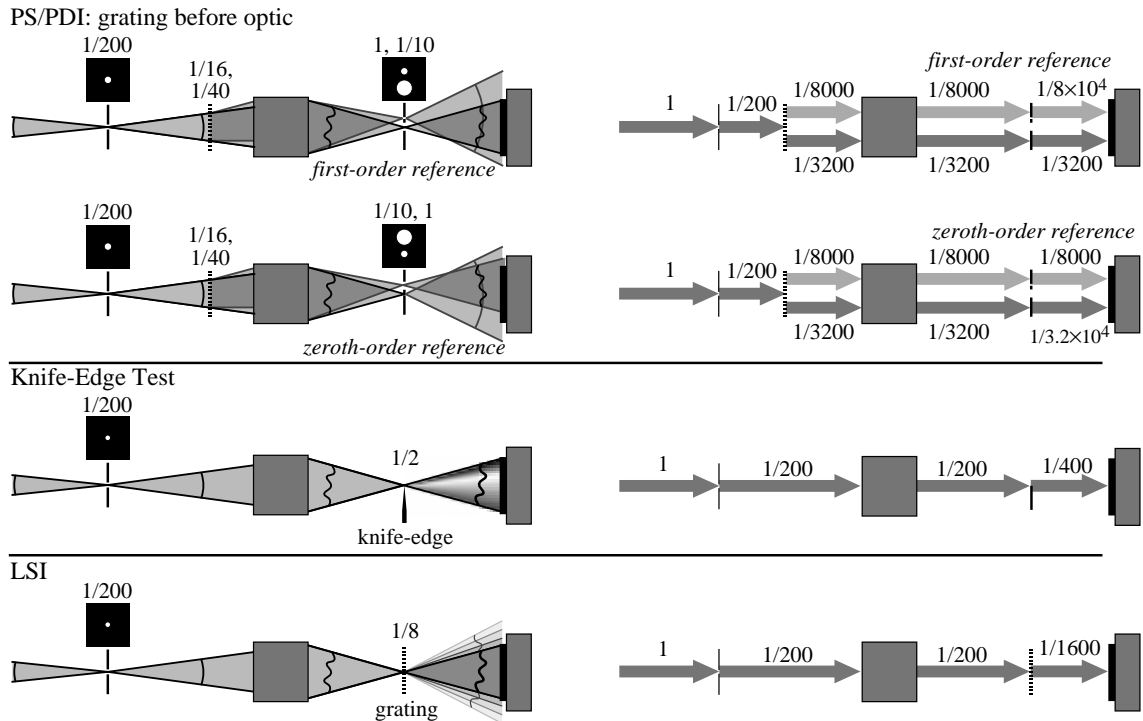
**Table 1.** Comparison of the relative intensities in six point diffraction interferometer configurations. The stationary intensity is the average intensity in the interference pattern, while the modulated intensity describes the half-height of the fringes. Contrast is the ratio of these two intensities. The efficiency comparison is based on the ratios of the modulated intensities among the different configurations shown. Intensity magnitudes are given relative to the illuminating beam upstream of the object pinhole, neglecting the efficiency of the test optic.

	Schematic	$I_{stationary}$ [ $\times 10^{-6}$ ]	$I_{modulated}$ [ $\times 10^{-6}$ ]	fringe contrast	relative efficiency
<b>Conventional PDI</b>					
a)		1	1	100%	$8 \times 10^{-3}$
<b>PS/PDI: grating after optic</b>					
b)		325	125	38%	1
c)		156	125	80%	1
<b>PS/PDI: grating before optic</b>					
d)		325	125	38%	1
e)		156	125	80%	1
<b>PS/PDI: upstream grating</b>					
f)		5625	5590	99%	45

given in Appendix 5. (Note: because the two interfering beams travel with the same divergence angles and fully overlap, the terms *intensity* and *flux* are used interchangeably in this discussion.) The stationary intensity represents the average amount of light recorded in the interference pattern, while the modulated intensity describes the intensity *height* of the interference fringes. 100% contrast is achieved when the two intensities have equal magnitude.

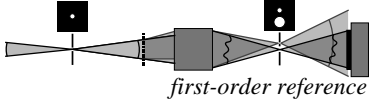
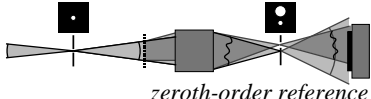
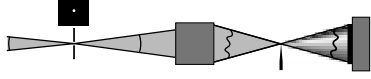
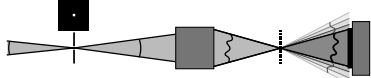
One significant advantage of the PS/PDI revealed in Table 1 is that the PS/PDI has over 100 times greater efficiency than the PDI. Comparison of the first-order and zeroth-order reference configurations produces two interesting results. First, the efficiency of the first-order reference configuration is twice as high because the brighter zeroth-order beam is unattenuated in the image-plane. Second, the modulated intensity is *the same* in the two configurations. This result is due to the fact that the modulated intensity comes from the cross-product of the two intensities, and is proportional to the geometric mean. No matter which of the two beams is attenuated by the spatial filtering, the geometric mean is the same. Having the same modulated intensity in the two configurations, the one with lower stationary intensity will produce higher fringe contrast — indeed, the contrast is twice as high in the zeroth-order reference configuration.

Another result of this comparison is the observation that the upstream grating configuration is 45 times more efficient than the configuration of the PS/PDI used for experiments. Because no object-plane spatial filtering is performed on the reference beam, to avoid beam overlap this configuration needs a very high-quality



**Figure 5.** A comparison of the efficiencies of the PS/PDI and two non-point-diffraction interferometer designs. (a) through (c) show the approximate efficiency of each element, neglecting the optic itself. Schematics (d) through (f) separately model the approximate integrated intensities of the test and reference beams as they propagate.

**Table 2.** Comparison of the relative intensities of three interferometer designs. The intensity magnitudes are given relative to the unfiltered illuminating beam, neglecting the efficiency of the test optic.

	Schematic	$I_{stationary}$ [ $\times 10^{-6}$ ]	$I_{modulated}$ [ $\times 10^{-6}$ ]	fringe contrast	relative efficiency
<b>PS/PDI</b>					
a)	 <i>first-order reference</i>	325	125	38%	1
b)	 <i>zeroth-order reference</i>	156	125	80%	1
<b>Knife-edge</b>					
c)		5000	5000	100%	40
<b>LSI</b>					
d)		1250	1250	100%	10

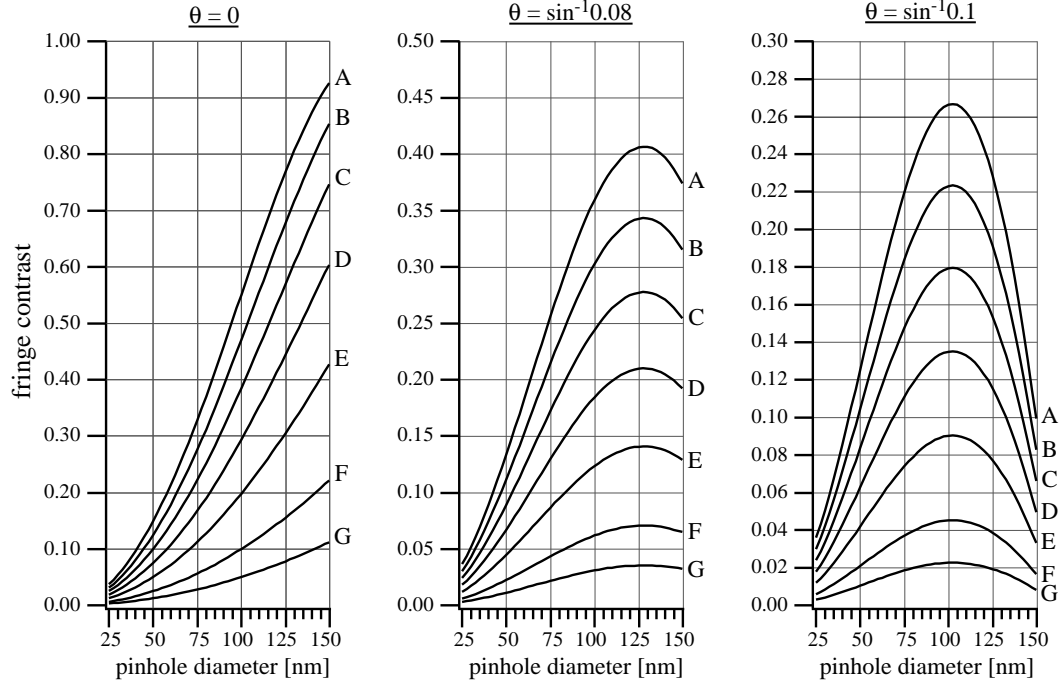
illuminating beam. If the reference beam were not of high quality, then the attenuation of the image-plane pinhole would be much greater and this configuration will have comparable efficiency to the others.

#### 4.4.1 Comparison with Other Interferometer Designs

In Fig. 5 and Table 2, an efficiency comparison is made between the PS/PDI, the Knife-Edge test, and the grating-based LSI. For a given application, it would appear that there is a necessary trade-off between efficiency and accuracy. When implemented experimentally, the efficiency advantages of the LSI design may be outweighed by the longer time required for analysis and larger uncertainties in the measurements. The time saved by the predicted factor-of-ten reduction in the single-image LSI exposure time may be undone by the increased analysis time and the need to record more exposures than in the PS/PDI scheme.

#### 4.6 CHOOSING THE OPTIMAL PINHOLE SIZE

Selecting the optimal pinhole diameter for a given application of the PS/PDI requires the balancing of several opposing concerns. The desire for a high degree of spatial filtering and a reference wave of uniform intensity motivates the use of the smallest available pinhole. However, the intensity of the reference wave is critical to achieving fringes of good contrast, a vital aspect of measurement precision. Based on a simple scalar diffraction model, this section outlines two methods for determining the optimal pinhole size for a given application, as applied to the study of EUV systems with 0.08 or 0.1 NA. Until such time as the results from a more detailed analysis of EUV pinhole diffraction (such as that presented in Chapter 2) are readily available, these two methods provide approximate results and illustrate important physical effects that require future study.



**Figure 6.** A simple model of the dependence of the fringe contrast on the pinhole diameter, based on “Airy pattern” diffraction from a circular reference pinhole at 13.4 nm wavelength. Given the relative strengths of the test and reference beams and the transmitting properties of the reference pinhole, the contrast dependence on pinhole diameter will follow one of the labeled contours from the graphs. For example, if a measured 100-nm reference pinhole produces 20% fringe contrast at the center of the interferogram, following contour “E” the fringe contrast will be 12% at the edge of the 0.08 NA and 9% at the edge of 0.1 NA. If a 50-nm pinhole were placed in this same system, then the expected fringe contrast would drop to 5.0% at the center, 4.4% at 0.08 NA and 4.2% at 0.1 NA.

#### 4.6.1 Reference Wave Uniformity

When the pinholes are smaller than the central lobe of the focal pattern of an optical system under test, the amplitude of the field transmitted through a pinhole should be roughly proportional to the pinhole area. Although this simple model neglects the complicated attenuating effects of high-aspect-ratio, highly absorptive pinholes on the order of a few wavelengths in diameter, it will serve as a good starting point for these calculations. To keep the model simple and useful, assume circular pinholes in opaque membranes, and scalar diffraction of ideal, Airy-like reference waves. With  $d$  as the pinhole diameter, the diffracted field amplitude  $E$  is

$$E(d, \theta) = Ad^2 \left[ 2 \frac{J_1\left(\frac{1}{2}kd \sin^{-1} \theta\right)}{\left(\frac{1}{2}kd \sin^{-1} \theta\right)} \right]. \quad (3)$$

$A$  is a constant multiplier dependent on the characteristics of the pinhole and on the relative strengths of the test and reference beams. If we define the amplitude of the test wave in the plane of the detector as 1, then  $A$  is on the order of  $1/d^2$ . Yet  $A$  is an experimental parameter and cannot be known ahead of time.

Using this simplified model, the intensity  $I$  of the interfering test and reference beams is

$$I = |e^{i\phi} + E|^2 = 1 + |E|^2 + 2|E|\cos\phi, \quad (4)$$

where  $\phi$  represents the phase of the test wavefront plus a significant PS/PDI spatial carrier frequency. The spatial carrier frequency typically introduces a large number of fringes, and the resultant field varies from its maximum to its minimum value over a short distance. The fringe contrast  $C$  is defined (Appendix 5) as

$$C = \frac{2|E|}{1 + |E|^2}. \quad (5)$$

When the intensities of the test and reference beams are matched, the contrast is one.

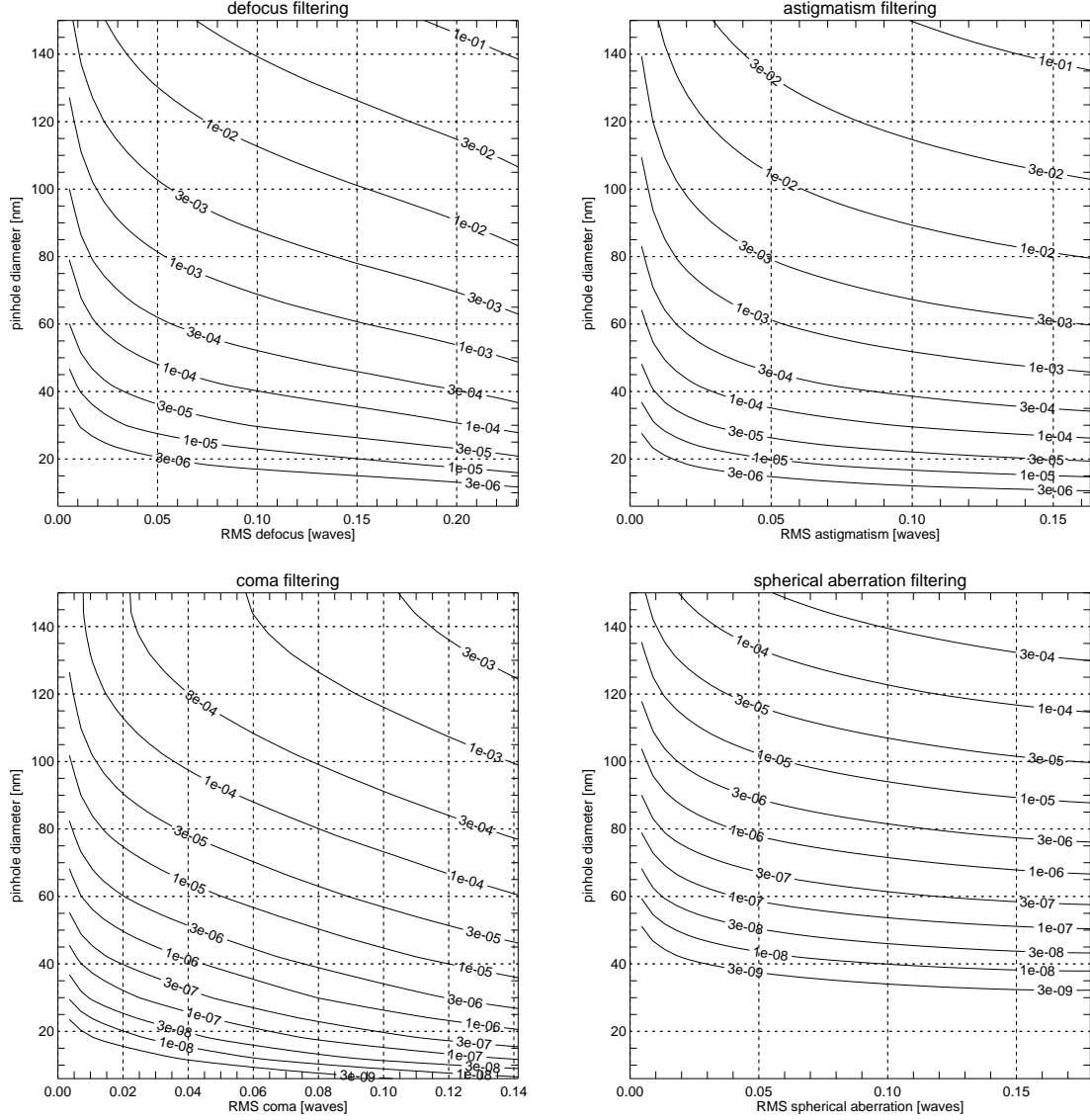
One goal in selecting a pinhole is to have high contrast across the entire NA of measurement. For several values of the parameter  $A$ , Fig. 6 shows the fringe contrast at the center of the interferogram and at the maximum polar angles within numerical apertures of 0.08 and 0.1. The contrast is calculated from Eqns. (3) and (5). The non-uniformity in the diffracted reference wave causes a greater contrast variation from the large pinholes than from the small pinholes. The corresponding labeled contours in the three graphs represent the same values of the parameter  $A$ .

#### 4.6.2 A Simple Approach to Pinhole Spatial Filtering Considerations

Determining the optimal reference pinhole size for a given PDI or PS/PDI application is a daunting task theoretically, and a laborious task experimentally. Abandoning the level of detail used in the TEM-PEST simulations of Chapter 2, a simple approach to this problem proves useful for assessing the level of spatial filtering produced by different pinhole sizes in the presence of aberrated test beams. Based only on Kirchoff diffraction from an idealized opaque planar screen, this study gives insight into the troublesome problems associated with filtering astigmatic aberrations.

In order to study the isolated effects of individual low-ordered aberrations, an initial 0.08 NA (reference) wavefront is given varying aberration magnitudes composed of a single low-ordered aberration component at a time. For this mathematical study (similar to studies by SangHun Lee), ideal circular pinholes of varying diameter are placed precisely at the center of the focal pattern produced by an optical system operating at 13.4 nm wavelength. In approximation to the Kirchoff boundary conditions, the simple discrete Fourier-transform (DFT) is used to mathematically propagate the scalar electric field (Sections 2.3 and 11.3.1). On propagation to the detector at far-field, the pinhole field produces the reference wavefront. A wavefront-phase analysis of the reference wave is performed within 0.08 NA, and the contributions of defocus, astigmatism, coma, and spherical aberration are identified. As the pinhole size is varied, the diffracted reference wavefront is studied within 0.08 NA. Displacement of the pinhole from the position of best-focus is not considered here.

This study is limited to the case where the pinhole is centered in the focal pattern. Experimentally,



**Figure 7.** A simple study of pinhole spatial filtering designed to assess the degree of filtering produced by different circular pinholes sizes in the presence of varying degrees of primary wavefront aberrations. Calculations are performed for a 0.08 NA optical system operating at 13.4 nm wavelength. Defocus, astigmatism, coma, and spherical aberration are investigated. The abscissa in each plot is given in RMS wavefront displacement of the single aberration component being investigated. The RMS wavefront displacement of the diffracted wave is given.

in the PS/PDI, every effort is made to center the pinhole in order to maximize the intensity transmitted into the reference wave. This situation is very different from the PDI, in which the pinhole is significantly displaced from the center of the pattern in order to produce an analyzable interference pattern.

Figure 7 contains the results of this study. Here, as the RMS aberration magnitudes are *increased*, the pinhole diameters required to produce a reference wavefront with an arbitrarily small RMS displacement (such as  $\lambda/100$ ) *decrease*. Of the four primary aberrations studied, astigmatism is by far the most troublesome, as it is the most difficult aberration to spatially filter. This property is born-out by the experimen-



tal observations that astigmatism creates the greatest uncertainty in the measurements (Chapters 3, 7, and 8). According to this simple model, in the presence of 0.1 waves RMS of astigmatism, the pinhole size required to filter the aberrations down to 0.03 waves ( $\lambda/33$ ) is 114 nm, to filter down to 0.01 waves ( $\lambda/100$ ) is 89 nm; and to filter down to 0.003 waves ( $\lambda/333$ ) is 67 nm.

This simple study leads to two important conclusions. First, the optimal pinhole size to achieve a desired reference wavefront quality depends strongly on the aberrations present in the system. Second, astigmatism is *the* most difficult aberration to filter. The measured astigmatism in sub-aperture A of the Schwarzschild objective examined in this thesis is 0.42 waves P-V or 0.0856 waves RMS (see Chapter 7). According to the simple calculation shown in Fig. 7, a sub-90-nm pinhole is required to filter this astigmatism magnitude to below 0.01 waves RMS in the reference beam. By comparison, coma and spherical aberration magnitudes much larger than this are easily filtered by considerably larger pinholes.

Because of its critical importance, more research in the area of EUV pinhole diffraction and spatial filtering is certainly required. Both detailed and simple calculations should support the experimental research so that a greater understanding of the pinhole size requirements of high-accuracy applications will be known. With the recent availability of pinholes from well-controlled pinhole fabrication processes at this small scale (fabricated by Erik Anderson) and the continued measurement of optical systems of various wavefront quality, important empirical data will be gathered.

# 5

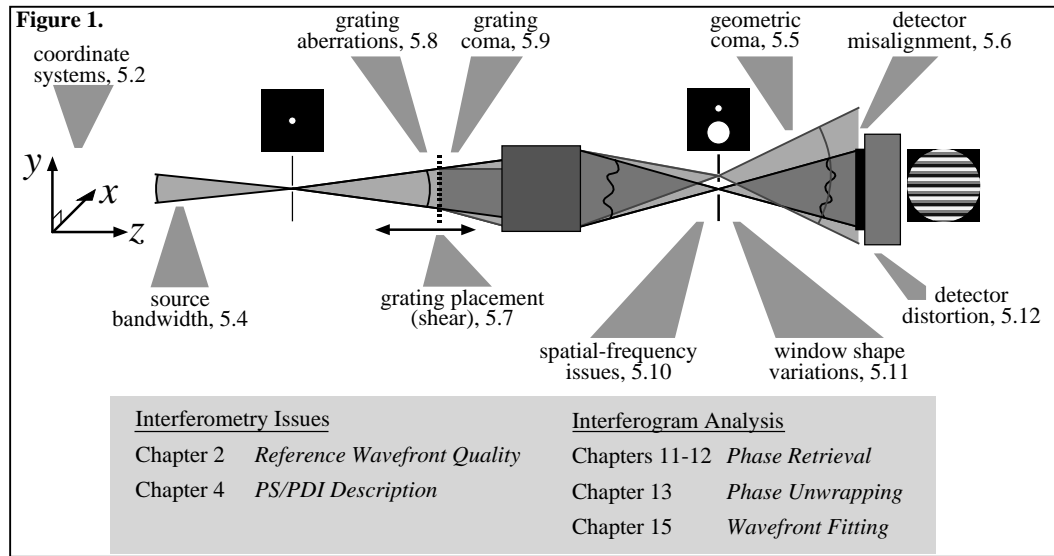
## *Systematic Errors and Measurement Issues*

---

<b>5.1 OVERVIEW</b>	<b>62</b>
5.1.1 Outline	
5.1.2 System Parameters	
<b>5.2 DEFINITION OF COORDINATE SYSTEMS</b>	<b>63</b>
5.2.1 The Laboratory System	
5.2.2 The Beam System	
5.2.3 The Detector System	
<b>5.3 NUMBER OF FRINGES</b>	<b>65</b>
5.3.1 Numbers	
<b>5.4 WHITE-LIGHT CONFIGURATION AND BANDWIDTH</b>	<b>68</b>
5.4.1 Effect of Bandwidth on the Measured Fringe Pattern	
5.4.2 Fringe Blurring in Symmetric Intensity Distributions	
5.4.3 Determining $W(d)$ for the Gaussian and Top-Hat Distributions	
5.4.3.1 Gaussian Distribution	
5.4.3.2 Top-Hat Distribution	
<b>5.5 GEOMETRICAL COMA SYSTEMATIC ERROR</b>	<b>74</b>
5.5.1 Representation of Zernike Pairs in Vector Notation	
5.5.2 Isolating and Removing the Geometric Coma Effect	
5.5.2.1 Method 1: Removing the Geometric Coma with Known Measurement NA	
5.5.2.2 Method 2: Removing Geometric Coma Using the Difference Wavefront	
5.5.2.3 Alternate Description of Method 2: Removing Geometric Coma	
<b>5.6 SYSTEMATIC ERROR FROM DETECTOR MISALIGNMENT</b>	<b>81</b>
5.6.6.1 Numbers	
<b>5.7 GRATING PLACEMENT CONSIDERATION: SHEAR</b>	<b>84</b>
<b>5.8 GRATING FABRICATION ERRORS</b>	<b>86</b>
5.8.1 Grating Aberrations	
5.8.2 Phase-Shifting	
5.8.3 Local Imperfections and Substrate Errors	
5.8.4 Recommendations	
<b>5.9 GRATING COMA</b>	<b>90</b>
<b>5.10 SPATIAL FILTERING BY THE IMAGE-PLANE WINDOW</b>	<b>94</b>
5.10.1 A Simple Model for Spatial Filtering	
5.10.2 Effect of Spatial Filtering on the Intensity and Phase Measurement	
5.10.3 Examples	
<b>5.11 VARIATIONS OF THE PS/PDI SPATIAL FILTER</b>	<b>98</b>
5.11.1 Image-Plane Window/Pinhole Filter Designs	
<b>5.12 DISTORTIONS DUE TO THE PLANAR DETECTOR</b>	<b>101</b>
<b>5.13 SUMMARY OF SYSTEMATIC ERRORS AND RECOMMENDATIONS</b>	<b>103</b>

## 5.1 OVERVIEW

In the pursuit of the highest achievable accuracy, it is important to consider all elements of the system, including the system geometry, as potential sources of systematic errors. This chapter is devoted to mathematical investigations of each of the PS/PDI components with the goal of identifying the most significant sources of systematic error. A very general approach is adopted so that this discussion may be applied to the design of interferometers for the measurement of arbitrary optical systems. Where appropriate, the results of these sections are applied to the specific configurations used in EUV interferometry of lithographic optics. These EUV calculations are highlighted at the end of each section and summarized at the end of this chapter. Random error sources and issues relating to inadequate pinhole spatial filtering are *not* covered in this chapter.



### 5.1.1 Outline

In low-NA configurations, a few definitions and formulae lead to simple mathematical descriptions of the various components. The predicted performance of an interferometer configuration can be judged based on the magnitude of the effects identified in this chapter. Figure 1 enumerates the most significant effects and indexes the sections of this thesis in which they are addressed.

### 5.1.2 System Parameters

There are at least three interferometer configurations of special interest here: both EUV and visible-

**Table 1:** Lithographic system parameters of interest. These numbers will be used for comparison throughout this chapter's investigation of systematic effects. Particular attention is paid to the EUV parameters as they pertain directly to experiments described in this thesis.

Parameter	EUV	Visible	Deep UV
wavelength, $\lambda$	13.4 nm	632.8 nm	193 nm
$NA_i$	~0.08	~0.08	~0.6-0.7
magnification	4-10	4-10	4-10
$\lambda/2NA$	.084 $\mu\text{m}$	3.96 $\mu\text{m}$	0.16 $\mu\text{m}$

light measurements of an EUV lithographic optic and, for comparison, a 193-nm-wavelength lithographic system with  $NA > 0.6$ . Approximate system parameters for each are given in Table 1.

## 5.2 DEFINITION OF COORDINATE SYSTEMS

Mathematical descriptions of the interferometer are simplified by the introduction of several inter-related coordinate systems, individually appropriate to different regions or components. This section introduces three coordinate systems and the expressions that relate them: the *Laboratory System*, the *Beam System*, and the *Detector System*.

Common to all of the coordinate systems is the NA of the beam in the region of interest, called the *local NA*. The local NA is determined by the sizes of various apertures and pupils in the system and describes the cone of rays that eventually reaches, or emanates from, an image or object point. These are the rays relevant to interferometric measurement of low spatial-frequency aberrations. In a reflective, cylindrically symmetric optical system,  $\alpha$  is defined as the maximum half-angle within the system NA. By definition,

$$NA \equiv \sin \alpha . \quad (1)$$

Where the spherical beams are incident on planar surfaces normal to the central ray, the tangent of  $\alpha$  is a useful quantity. Define  $t$  as

$$t \equiv \tan \alpha = \frac{NA}{\sqrt{1 - NA^2}} . \quad (2)$$

### 5.2.1 The Laboratory System

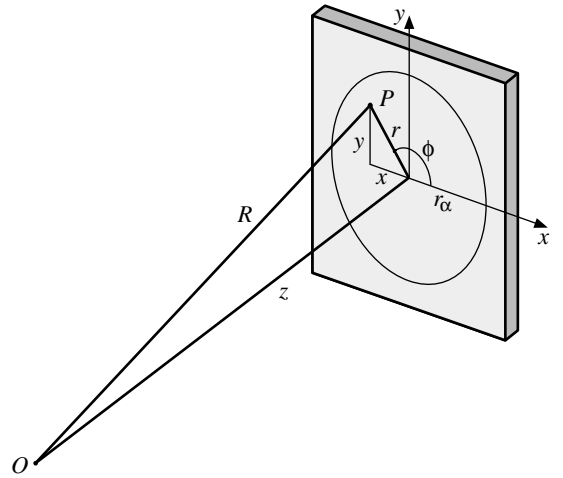
The Cartesian system, or *Laboratory Coordinates*, shown in Fig. 2, defines points in 3-D space as  $P(x, y, z)$ . The  $z$ -axis coincides with the central ray of the interferometer's *test* beam; the origin of the coordinates is defined as the center of curvature of the diverging (or converging) spherical beam. This point is typically determined by the position of a pinhole spatial filter in either the object- or the image-plane, or by the focal point of the interferometer's test beam.

$$R = \sqrt{x^2 + y^2 + z^2} . \quad (3)$$

$$r = \sqrt{x^2 + y^2} . \quad (4)$$

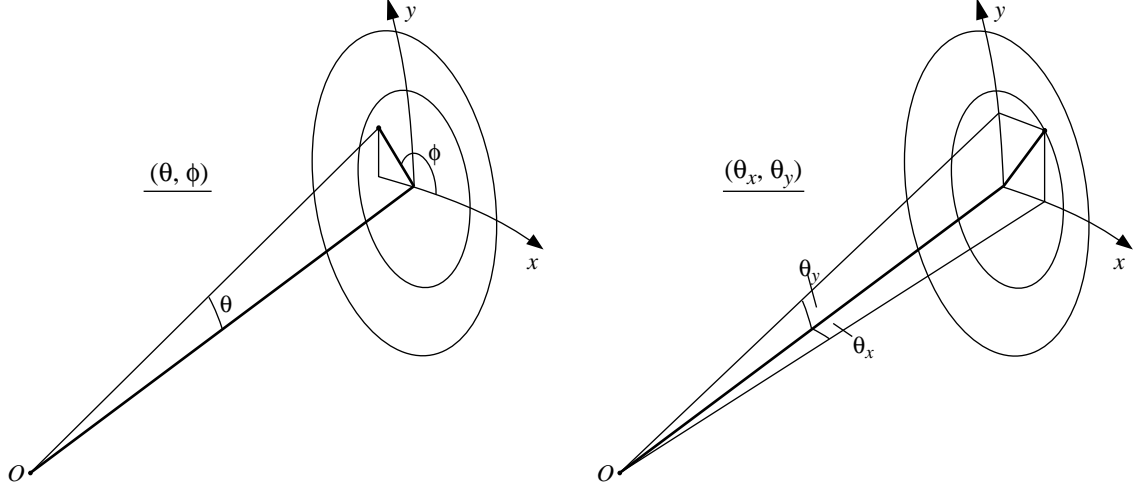
Where spherical beams are incident on a planar surface, cylindrical coordinate systems have a maxi-

The Laboratory Coordinates



**Figure 2.** The *Laboratory Coordinate System* is based on Cartesian or cylindrical coordinates using the measurement units of the experimental system. The  $z$ -axis is defined to be collinear with the central ray of the *test* beam, with the center-of-curvature as the origin.

Two equivalent representations of the spherical *Beam Coordinates*



**Figure 3.** The spherical *Beam Coordinate System* uses polar and azimuthal angles to represent an *angular position* within the diverging or converging spherical beam. An alternate representation of the same coordinate system defines an angular position using an  $x$  and  $y$  pair of polar angles. Again, the  $z$ -axis is defined to be collinear with the central ray of the *test* beam, with the center of beam curvature as the origin.

mum radius,  $r_\alpha$  corresponding to rays at the angle  $\alpha$  from the axis.

$$r_\alpha \equiv z \tan \alpha = z t . \quad (5)$$

### 5.2.2 The Beam System

A spherical coordinate system, shown in Fig. 3, provides a more natural description of the diverging or converging beams: for the optical systems of interest here, aberrations are described as departures from an ideal, spherical wavefront. Based on the central ray of the test beam, we define a *position* within the beam using the polar and azimuthal angles  $(\theta, \phi)$ . It will also be convenient to define a polar angle *vector*, separated into  $x$  and  $y$  angular components:

$$\boldsymbol{\theta} \equiv (\theta_x, \theta_y) \equiv (\theta \cos \phi, \theta \sin \phi). \quad (6)$$

In some cases, this angular vector simplifies translation to the Cartesian Laboratory System. Other expressions in this coordinate system relate  $\boldsymbol{\theta}$  to  $\mathbf{k}$ , which is also used to represent the beam propagation direction:,

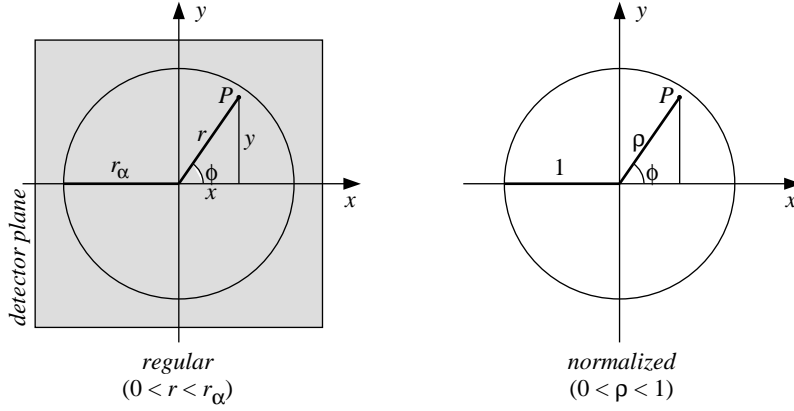
$$\mathbf{k} = (k_x, k_y, k_z) = (\sin \theta \cos \phi, \sin \theta \sin \phi, \cos \theta), \quad |\mathbf{k}| = 1 . \quad (7)$$

$$\theta_{x,y} = \tan^{-1} \left( \frac{k_{x,y}}{k_z} \right) . \quad (8)$$

Normalization of the polar angle relative to the local NA will simplify calculations in some cases. This system is called the *Normalized Beam Coordinates*. For this purpose, define a normalized *angle*  $\gamma$  as

$$\gamma \equiv \frac{\theta}{\alpha} . \quad (9)$$

The Detector Coordinates



**Figure 4..** The *Detector Coordinate System* is used to represent measurements recorded on a planar detector. Both Cartesian and polar representations are used. The origin is taken to be the point of intersection of the central ray of the test beam and the detector plane.

### 5.2.3 The Detector System

The final coordinate system introduced here is the 2-D polar *Detector Coordinates*, defined in the plane of the detector and centered on the point of intersection of the central ray of the test beam with the detector plane. The Detector Coordinate System is shown in Fig. 4. A point in the detector plane may be represented in the units of the Laboratory frame, or by a corresponding set of *Normalized Detector Coordinates* utilizing a dimensionless radius  $\rho$  based on  $r_\alpha$ .

$$\rho \equiv \frac{r}{r_\alpha} . \quad (10)$$

From Eqns. (2) and (10) we also have the relationships,

$$r = r_\alpha \rho = t z \rho . \quad (11)$$

The normalized coordinates  $(\rho, \phi)$  eventually become the coordinate system of the data analysis, which is based on a unit circle representation of the system's exit pupil.

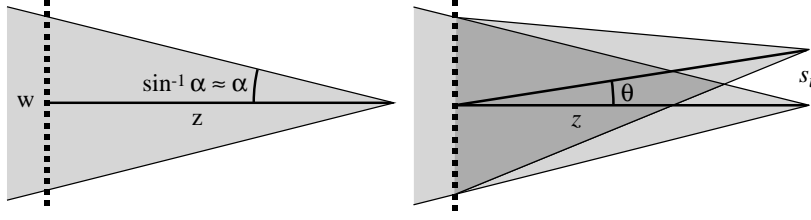
## 5.3 NUMBER OF FRINGES

From the mathematical description of the PS/PDI arises a convenient rule of thumb useful for determining the required position and pitch of a grating beamsplitter.

**Rule of Thumb:** The number of fringes in the interferogram equals the number of grating lines illuminated within the NA of the optical system.

Subject to the fact that wavefront aberrations in the test optic cause curvature in the observed fringe patterns, this rule is approximate. An investigation of the origin of this rule leads to a description of how the PS/PDI can be used with broadband illumination (Section 5.4).

Let  $\alpha_i$  and  $\alpha_o$  be the maximum half-angles within the image-side and object-side numerical apertures  $NA_i$  and  $NA_o$  respectively. Constraining our discussion to one side of the optical system, in general



**Figure 5.** A grating beam-splitter is used to produce the test and reference beams of the PS/PDI. The first-order diffraction angle  $\theta$  is determined only by the grating pitch  $d$ . The beam separation in the image-plane  $s_i$  depends on  $\theta$  and on the position  $z$  of the grating with respect to focus.

we have,

$$NA = \sin \alpha, \quad \text{and} \quad t \equiv \tan \alpha = \frac{NA}{\sqrt{1 - NA^2}}. \quad (12) \text{ and } (13)$$

Depending on the configuration,  $z$  is defined as the distance from the grating to the object-plane or as the distance from the grating to the image-plane.

For either an object-side or image-side grating (Chapter 4), the lateral width of the grating illuminated within the local NA is  $w$ , as shown in Fig 5.

$$w = 2zt. \quad (14)$$

Therefore for a given grating pitch  $d$  the number of grating lines illuminated is

$$N_{lines} = \frac{w}{d} = \frac{2zt}{d}. \quad (15)$$

When the grating is placed on the image-side, the converging beam from the optical system forms a series of real images corresponding to the diffraction orders of the grating. The lateral separation of two adjacent image-side foci  $s_i$  follows from the grating equation for the first diffracted order  $\lambda = d\theta$  where  $\theta$  is typically much smaller than  $\alpha$ .

$$s_i \approx z\theta \approx \frac{z\lambda}{d}. \quad (16)$$

Given  $s_i$ , the number of fringes within the NA is readily calculated from the maximum path length difference between the zeroth and the first diffracted orders. By symmetry, this maximum difference  $\Delta$  is twice the difference between the central ray, and the rays at the angle  $\alpha_i$ .

$$\Delta_{1/2} \approx s_i \sin \alpha_i = s_i NA_i, \quad \text{and} \quad \Delta = 2\Delta_{1/2} \approx \frac{2z\lambda}{d} NA_i. \quad (17) \text{ and } (18)$$

The number of fringes  $N_{fringes}$  is equal to the path length difference in waves (units of  $\lambda$ ).

$$N_{fringes} = \frac{\Delta}{\lambda} \approx \frac{2z}{d} NA_i \approx \frac{2zt_i}{d} = N_{lines}, \quad (19)$$

thereby justifying the rule of thumb. This number may also be written using  $s_i$  explicitly

$$N_{fringes} \approx N_{lines} \approx \frac{2s_i t_i}{\lambda}. \quad (20)$$

When the grating is placed on the object-side, the rule of thumb still applies. In this case, however,

the grating divides the *diverging* beam, and each grating order besides the zeroth appears to originate from a separate virtual object source. By analogy with Eq. (16), the separation of these virtual objects is

$$s_o \approx z\theta \approx \frac{z\lambda}{d}. \quad (21)$$

In principle, the separation of the real foci in the image-plane is equal to the separation of the virtual objects, scaled by the magnification of the system.

$$s_i \approx ms_o. \quad (22)$$

The relationship of the object-side and image-side NA angles,

$$\alpha_i = \alpha_o / m, \quad (23)$$

allows Eq. (20) to be written independent of the system magnification.

$$N_{\text{fringes}} \approx N_{\text{lines}} \approx \frac{2s_i t_i}{\lambda} \approx \frac{2s_o t_o}{\lambda}. \quad (24)$$

In the small-NA range where  $\sin \alpha \approx \tan \alpha$  a useful approximation for the number of fringes is

$$N_{\text{fringes}} \approx N_{\text{lines}} \approx \frac{2s_i}{\lambda} NA_i \approx \frac{2s_o}{\lambda} NA_o. \quad (25)$$

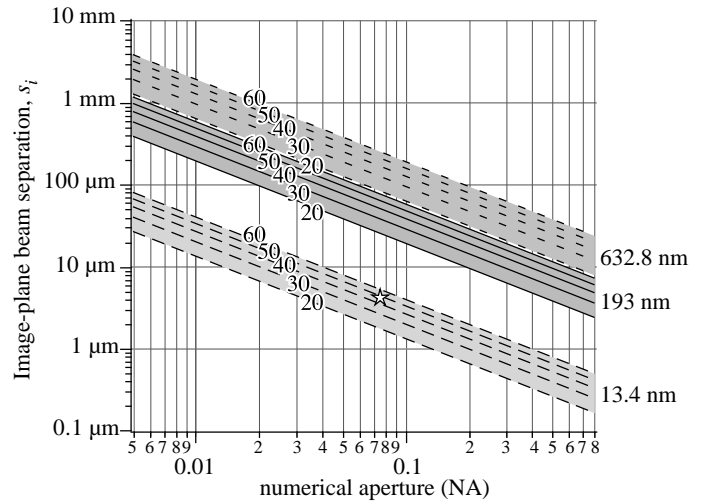
In the *upstream grating configuration* (Section 4.3) the multiple object sources are real, and the same rules apply.

### 5.3.1 Numbers

For a specified number of fringes, we can investigate the corresponding image-plane beam separation by solving for  $s_i$  in Equation (25):

$$s_i = \frac{N\lambda}{2NA_i}. \quad (26)$$

Figure 6 shows Eq. (26) plotted versus NA for three wavelengths of interest. To achieve 40 fringes at 0.08 NA in the EUV, 193-nm, and HeNe configurations requires beam separations of 4.2  $\mu\text{m}$ , 60  $\mu\text{m}$ , and 198  $\mu\text{m}$  respectively. Forty fringes at 193 nm with 0.6 NA requires a beam separation of 6.4  $\mu\text{m}$ .



**Figure 6.** The number of interferogram fringes depends on the wavelength, the image-plane beam separation, and the measurement NA. For a wide range of numerical apertures, this figure shows the beam separation required to produce a given number of fringes. Three experimentally relevant wavelengths are considered: EUV (13.4 nm), deep UV (193 nm) and visible (632.8, HeNe). The star indicates EUV numbers relevant to experiments conducted in this thesis.



### 5.3 Summary

**Beam Separation.**  $s/N_{fringe} = \lambda/2t = 0.084 \mu\text{m/fringe} \Rightarrow N_{fringe}/s = 2t/\lambda = 12 \text{ fringes}/\mu\text{m}$ .

### 5.4 WHITE-LIGHT CONFIGURATION AND BANDWIDTH

For the grating-based configurations of the PS/PDI, the rule of thumb presented in the previous chapter equates the number of grating lines illuminated to the number of fringes observed. The fact that this rule is independent of the illumination wavelength leads to the conclusion that, aside from chromatic aberrations (wavelength-dependent effects) in the test optical system, the PS/PDI may be regarded as a *broad-band* interferometer. This section describes the most important wavelength-dependent effects of the interferometer. Note that this discussion addresses only an ideal, diffraction-limited, achromatic optical system under test.

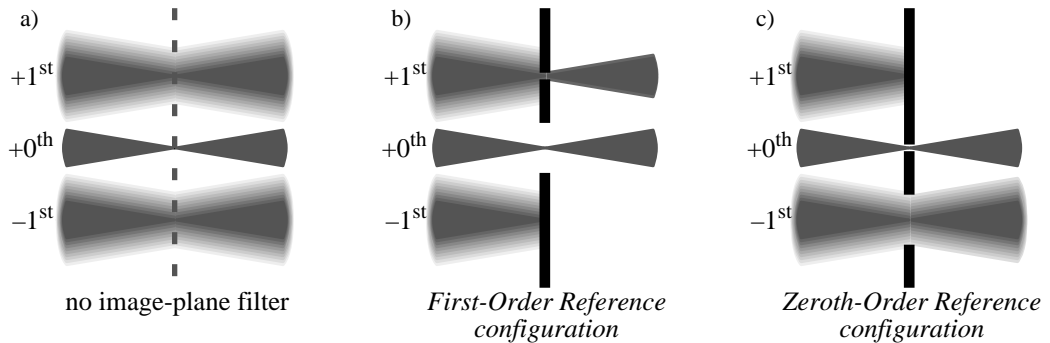
Since the NA is a property of an optical system independent of wavelength, the number of grating lines that fall within the NA is determined only by the geometry. For a given wavelength, the number of observed fringes is related to the image-plane beam separation  $s_i$ , according to Eq. (24):.

$$N_{fringes} \approx N_{lines} \approx \frac{2s_i t_i}{\lambda}. \quad (27)$$

Using  $N$  as a convenient system invariant,  $s_i$  may be written as

$$s_i \approx \frac{N\lambda}{2t_i}, \quad (28)$$

showing that the image-plane beam separation is proportional to the wavelength. When the illumination is not monochromatic, grating-diffracted beams are focused to different lateral positions in the image-plane. The position of the zeroth-order focus does not depend on the grating pitch, and is thus not wavelength-dependent.



**Figure 7.** When a grating is used to separate the test beam, the diffractive orders are affected by the bandwidth of the illumination. (a) Different wavelength components of the first-order beams are separated by a lateral displacement in the image plane. (b) In the first-order reference configuration, the reference pinhole behaves as a monochromator, selectively transmitting a portion of the bandwidth more effectively than the rest. (c) A small translation of the image-plane spatial filter puts the system into the zeroth-order reference configuration, in which a much broader range of wavelengths is transmitted.

In the image-plane, the positive and negative first-order beams form foci on opposite sides of the zeroth-order beam. As shown in Fig. 7, a simple lateral translation of the two-pinhole spatial filter allows selection of either of the two first-order beams or of the zeroth-order as the reference or test beam. These two configurations are referred to as the *first-order reference* and the *zeroth-order reference* configurations, respectively. These names indicate which beam is filtered by the reference pinhole. In the presence of finite-bandwidth illumination, these two configurations do behave somewhat differently. Some advantages and disadvantages of these two similar arrangements are discussed in Section 5.8.

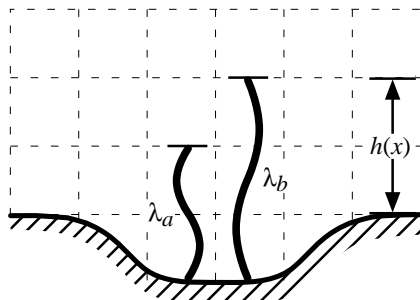
In the first-order reference configuration, the tiny reference pinhole serves as a monochromator by geometrically selecting some wavelengths to pass through the pinhole more efficiently than others. For this to be true, however, the test optic must be of nearly diffraction-limited quality. In the complementary zeroth-order reference configuration, the large window transmits a range of wavelengths determined by the window size and position. Hence, the range of recorded wavelength components may be different in the two configurations.

Due to the typically long time scale of measurement, relative to the frequency of the radiation, light of different wavelengths adds *incoherently*. Therefore, if there are wavelength components present in either the test or the reference beam but not in both, those components will contribute only to the unmodulated background intensity in the recorded data. *Unmodulated* refers to the recorded light that does not contribute to the interference fringes.

#### 5.4.1 Effect of Bandwidth on the Measured Fringe Pattern

When using the PS/PDI with broad-band illumination, interpretation of the measured interference fringe patterns may require careful consideration. Chromatic aberrations and geometrical considerations must both be considered. The relevant bandwidth here is *not* the source bandwidth; strictly speaking, it is only those wavelength components that reach the detector and are present in both the test and reference beams. These are the only wavelength components capable of producing interference fringes. For reasons stated above, this restriction may exclude some components of the original source bandwidth.

One design consideration of the interferometer is that different wavelength components separated



**Figure 8.** The origin of interferogram chromatic-dependence in a reflective achromatic system. For a particular wavelength component, the interferogram fringes reveal the optical path difference between two waves, measured in wavelengths. With a mirror surface height profile  $h(x)$  different fringe patterns will be observed for each wavelength of measurement. The surface depression shown in the figure is one-half of  $\lambda_a$  yet is one-third of  $\lambda_b$ . Upon reflection the aberration path length is doubled. Thus, these two wavelengths generate different fringe patterns for the same aberration.

by a grating placed before (on the object-side of) the optical system will travel along different paths through the optical system. The significance of this effect must be evaluated based on the illumination bandwidth and the design on the test optical system.

If a range of wavelengths is present, then the measured interferogram will be an additive combination of the wavelength intensity components allowed to reach the detector. In a perfect optical system, the pattern of equally-spaced, parallel interference fringes is the same for *all wavelengths*. However, in the presence of aberrations, each wavelength component may contribute a different, overlapping interference pattern, thereby confusing measurement.

Even in the absence of true chromatic aberrations, geometrical effects can contribute wavelength-dependence to measurements. For example, consider a reflective optical system with a surface figure error of arbitrary depth (or height). Light reflected from the region of a depression travels a relatively longer distance than the light in adjacent areas. This situation is depicted in Fig. 8.

The significance of a given path-length difference on the interference pattern is inherently wavelength-dependent. For each spectral component, the interference fringes represent contours of constant path-length difference, separated by one wavelength. Thus, for a given path-length difference, different wavelengths will generate different fringe patterns. In the presence of finite bandwidth illumination, this effect can *blur* a fringe pattern. However, a special situation arises if the spectral intensity distribution is symmetric about a central wavelength: the contrast is reduced, but the fringe positions are unaffected. Such a situation only affects the signal-to-noise ratio of the wavefront measurements. This can be demonstrated mathematically as follows.

The measured intensity is the sum of the intensity contributions from all of the available wavelengths. Eq. (3) defines an intensity-weighting function  $w(\lambda)$  with units of  $\lambda^{-1}$ , and an intensity function  $J(\mathbf{r};\lambda)$ .

$$I(\mathbf{r}) = \int_{\lambda} dI(\mathbf{r};\lambda) = \int_{\lambda} w(\lambda)J(\mathbf{r};\lambda)d\lambda, \quad (29)$$

$$\text{where} \quad \int_{\lambda} w(\lambda)d\lambda = 1. \quad (30)$$

A general expression for a single wavelength component  $J(\mathbf{r};\lambda)$  is

$$J(\mathbf{r};\lambda) = A + B \cos \left[ \mathbf{\kappa} \cdot \mathbf{r} + \frac{4\pi h(\mathbf{r})}{\lambda} \right], \quad (31)$$

where the vector  $\mathbf{\kappa}$  represents the spatial carrier frequency of the fringe pattern, and is invariant of wavelength.  $h(\mathbf{r})$  is the combined mirror figure *error* as seen by a given ray in an arbitrary reflective optical system. The path length of a particular ray is doubled upon reflection from a surface, as the light must twice travel the distance  $h$ .

Consider the addition of two different closely-spaced wavelength components  $\lambda_o - \Delta\lambda$  and  $\lambda_o + \Delta\lambda$  with  $\Delta\lambda \ll \lambda_o$ .

$$\frac{1}{2} J(\mathbf{r}; \lambda_o - \Delta\lambda) + \frac{1}{2} J(\mathbf{r}; \lambda_o + \Delta\lambda) = A(\mathbf{r}) + \frac{1}{2} B(\mathbf{r}) \cos \left[ \kappa \cdot \mathbf{r} + \frac{4\pi h(\mathbf{r})}{\lambda_o - \Delta\lambda} \right] + \frac{1}{2} B(\mathbf{r}) \cos \left[ \kappa \cdot \mathbf{r} + \frac{4\pi h(\mathbf{r})}{\lambda_o + \Delta\lambda} \right] \quad (32a)$$

$$\approx A(\mathbf{r}) + B(\mathbf{r}) \cos \left[ \kappa \cdot \mathbf{r} + \frac{4\pi h(\mathbf{r})}{\lambda_o} \right] \cos \left[ \frac{4\pi h(\mathbf{r})}{\lambda_o} \frac{\Delta\lambda}{\lambda_o} \right] \quad (32b)$$

$$\approx A(\mathbf{r}) + B(\mathbf{r}) \cos \left[ \kappa \cdot \mathbf{r} + \frac{4\pi h(\mathbf{r})}{\lambda_o} \right] \left\{ 1 - \frac{\pi^2}{8} \left[ \frac{h(\mathbf{r})}{\lambda_o} \right]^2 \left( \frac{\Delta\lambda}{\lambda_o} \right)^2 \right\}. \quad (32c)$$

In Eq. (32b), the wavelength-dependence is expanded to first-order in the  $\Delta\lambda$ . In the limit of narrow bandwidth  $\Delta\lambda$  or small aberrations  $h(\mathbf{r})$  Eqns. (32a) to (32c) *reduce* to the intensity pattern of the central wavelength, as expected. This fact leads to the illustration of an important result, worth elaboration.

The addition of two closely spaced and equally-weighted wavelength components yields the same fringe pattern as the central wavelength, with a reduction of fringe modulation determined by the magnitude of  $h(\mathbf{r})$  and  $\Delta\lambda$ .

#### 5.4.2 Fringe Blurring in Symmetric Intensity Distributions

In the special case of symmetric intensity distributions we can derive a general form of the resultant fringe pattern. The predicted reduction of the fringe modulation can be used as a criterion to set an upper limit on the allowable bandwidth. Non-symmetric distributions may be represented by the addition of a symmetric distribution with an asymmetric distribution. In this case, the following treatment would apply to the symmetric part, and the asymmetric part would have to be addressed separately.

When the wavelength distribution  $w(\lambda)$  is symmetric about  $\lambda_o$ , pairs of intensity components within the distribution add to re-create the pattern of the central wavelength. For a symmetric distribution,

$$w(\lambda_o - \Delta\lambda) \equiv w(\lambda_o + \Delta\lambda). \quad (33)$$

The sum of a pair of intensities within the distribution integral of Eq. (29) will be

$$\begin{aligned} & w(\lambda_o - \Delta\lambda) J(\mathbf{r}; \lambda_o - \Delta\lambda) + w(\lambda_o + \Delta\lambda) J(\mathbf{r}; \lambda_o + \Delta\lambda) \\ & \approx 2w(\lambda_o + \Delta\lambda) \left( A(\mathbf{r}) + B(\mathbf{r}) \cos \left[ \kappa \cdot \mathbf{r} + \frac{4\pi h(\mathbf{r})}{\lambda_o} \right] \left\{ 1 - \frac{\pi^2}{8} \left[ \frac{h(\mathbf{r})}{\lambda_o} \right]^2 \left( \frac{\Delta\lambda}{\lambda_o} \right)^2 \right\} \right). \end{aligned} \quad (34)$$

By symmetry, and Eq. (34), we have

$$I(\mathbf{r}) = \int_{\lambda} w(\lambda) J(\mathbf{r}; \lambda) d\lambda = \int_{\Delta\lambda} w(\lambda_o + \Delta\lambda) J(\mathbf{r}; \lambda_o + \Delta\lambda) d(\Delta\lambda) \quad (35a)$$

$$\approx 2 \int_{\Delta\lambda > 0} w(\lambda_o + \Delta\lambda) \left( A(\mathbf{r}) + B(\mathbf{r}) \cos \left[ \kappa \cdot \mathbf{r} + \frac{4\pi h(\mathbf{r})}{\lambda_o} \right] \left\{ 1 - \frac{\pi^2}{8} \left[ \frac{h(\mathbf{r})}{\lambda_o} \right]^2 \left( \frac{\Delta\lambda}{\lambda_o} \right)^2 \right\} \right) d(\Delta\lambda) \quad (35b)$$

$$\approx A(\mathbf{r}) + B(\mathbf{r}) \cos \left[ \mathbf{k} \cdot \mathbf{r} + \frac{4\pi h(\mathbf{r})}{\lambda_o} \right] \left\{ 1 - \frac{\pi^2}{8} \left[ \frac{h(\mathbf{r})}{\lambda_o} \right]^2 \int_{\Delta\lambda} w(\lambda_o + \Delta\lambda) \left( \frac{\Delta\lambda}{\lambda_o} \right)^2 d(\Delta\lambda) \right\} \quad (35c)$$

This representation allows us to define a bandwidth-dependent *fringe blurring function*  $W(\delta)$  to simplify this discussion,

$$W(\delta) \equiv \frac{\pi^2}{8} \int_{\Delta\lambda} w(\lambda_o + \Delta\lambda) \left( \frac{\Delta\lambda}{\lambda_o} \right)^2 d(\Delta\lambda). \quad (36)$$

The dimensionless parameter  $\delta$  that describes the *width* of the spectral intensity distribution will be defined differently for different distribution models (Gaussian, top-hat, etc.). Continuing to simplify Eq. (35c),

$$I(\mathbf{r}) \approx A(\mathbf{r}) + B(\mathbf{r}) \cos \left[ \mathbf{k} \cdot \mathbf{r} + \frac{4\pi h(\mathbf{r})}{\lambda_o} \right] \left\{ 1 - \left[ \frac{h(\mathbf{r})}{\lambda_o} \right]^2 W(\delta) \right\}, \quad (37a)$$

$$I(\mathbf{r}) \approx J(\mathbf{r}; \lambda_o) - B(\mathbf{r}) \cos \left[ \mathbf{k} \cdot \mathbf{r} + \frac{4\pi h(\mathbf{r})}{\lambda_o} \right] \left[ \frac{h(\mathbf{r})}{\lambda_o} \right]^2 W(\delta). \quad (37b)$$

In Eqns. (37a) and (37b), the bandwidth-dependent term acts to reduce the magnitude of the fringe modulation *B* without changing the positions of the inflection points. Furthermore, when  $A(\mathbf{r})$  and  $B(\mathbf{r})$  are slowly varying functions relative to the spatial period of the fringes, then the positions of the fringe maxima and minima will match the monochromatic case. Thus in the presence of a symmetric spectral distribution, the fringe contrast is merely reduced, and  $W(\delta)$  represents the fractional loss in fringe modulation.

#### 5.4.3 Determining $W(\delta)$ for the Gaussian and Top-Hat Distributions

For quantitative results we investigate two spectral intensity distribution models: *Gaussian*, and *top-hat*. The Gaussian distribution is defined by its full-width at half-maximum (FWHM), defined as  $\delta_g$ ; the top-hat is defined simply by its full-width  $\delta_t$ .

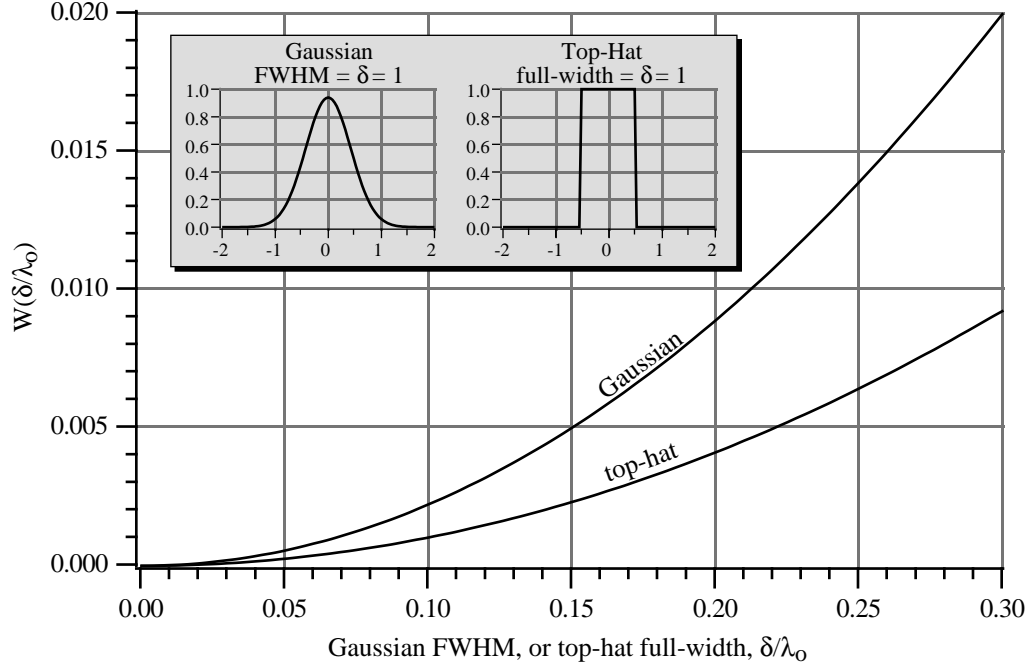
**5.4.3.1 Gaussian Distribution.** Consider a Gaussian distribution centered about  $\lambda_o$ , with an RMS width of  $\lambda_o \sigma$ . In this definition,  $\sigma$  is the dimensionless parameter describing the distribution width relative to the central wavelength. Normalization requires that Eq. (30) must be satisfied.

$$w_g(\lambda_o + \Delta\lambda) = \frac{1}{\sqrt{2\pi}\sigma} \exp \left[ -\frac{(\Delta\lambda)^2}{2\sigma^2} \right]. \quad (38)$$

Solution of the integral in Eq. (36) yields  $W_g(\sigma)$ .

$$W_g(\sigma) = \frac{\pi^2 \sigma^2}{8}. \quad (39)$$

As stated previously,  $W_g$  is defined differently for each distribution shape; the definition may use any convenient parameter that describes the distribution width. Often, a more convenient representation of the



**Figure 9.** A plot of the fringe-blurring function  $W(\delta)$  for the Gaussian and top-hat spectral distributions as a function of the distribution widths.  $W(\delta)$  is a parameter that describes the reduction in fringe height that can be expected in PS/PDI interferometry in the presence of a symmetric spectral distribution. Note that the two distribution widths are defined differently: this primarily accounts for the difference in  $W(\delta)$ . The two width definitions are illustrated in the inset graphs.

**Table 2.** Values of the fringe blurring function  $W(\delta)$  for the Gaussian and the top-hat spectral intensity distributions, and a selected distribution widths,  $\delta$ .  $W$  describes the expected reduction in fringe contrast related to non-monochromatic illumination of a reflective optical system. \* is the measured bandwidth used in EUV interferometry.

$\delta$	$W_g(\text{Gaussian})$	$W_t(\text{top-hat})$
0.000	0.00	0.00
0.001*	$2.22 \times 10^{-7}$	$1.03 \times 10^{-7}$
0.002	$8.90 \times 10^{-7}$	$4.11 \times 10^{-7}$
0.005	$5.56 \times 10^{-6}$	$2.57 \times 10^{-6}$
0.010	$2.22 \times 10^{-5}$	$1.03 \times 10^{-5}$
0.020	$8.90 \times 10^{-5}$	$4.11 \times 10^{-5}$
0.050	$5.56 \times 10^{-4}$	$2.57 \times 10^{-4}$
0.100	$2.22 \times 10^{-3}$	$1.03 \times 10^{-3}$
0.200	$8.90 \times 10^{-3}$	$4.11 \times 10^{-3}$

Gaussian distribution will be in terms of the full-width at half-maximum (FWHM)  $\delta_g$  rather than  $\sigma$ .

$\delta_g$  and  $\sigma$  are related by a constant coefficient. It is easily shown that

$$(\text{FWHM})^2 \equiv \delta_g^2 = (8 \ln 2) \sigma^2. \quad (40)$$

Defining  $W_g$  using the FWHM  $\delta_g$  instead of the RMS width  $\sigma$ ,

$$W_g(\delta_g) = \frac{\pi^2 \sigma^2}{8} = \frac{\pi^2 \delta_g^2}{64 \ln 2} = 0.222 \delta_g^2. \quad (41)$$

**5.4.3.2 Top-Hat Distribution.** In a similar manner as above, choose a normalized top-hat distribu-

tion centered about  $\lambda_o$  with a full-width  $\lambda_o\delta$ .

$$w_t(\lambda_o + \Delta\lambda) = \begin{cases} \frac{1}{\lambda_o\delta_t}, & |\Delta\lambda| < \frac{1}{2}\lambda_o\delta_t \\ 0, & |\Delta\lambda| > \frac{1}{2}\lambda_o\delta_t \end{cases}. \quad (42)$$

From Eq. (36) the fringe-blurring function is

$$W_t(\delta_t) = \frac{\pi^2\delta_t^2}{96} = 0.103\delta_t^2. \quad (43)$$

Selected values of the blurring-function  $W(\delta)$  from Fig. 9 are listed in Table 2. Even for significant bandwidths, the magnitude of the blurring-function, shows that in the presence of mirror surface aberrations, small on the scale of the central wavelength, the fringe modulation is not substantially reduced.

#### 5.4 Summary

**Bandwidth.**  $W_g = 2.22 \times 10^{-7}$  @ 0.1% BW (Gaussian distribution). Fringe amplitude is reduced by  $2.22 \times 10^{-7}$  per wave<sup>2</sup> of aberration at this bandwidth.

### 5.5 GEOMETRICAL COMA SYSTEMATIC ERROR

This section describes a systematic, geometric coma error introduced by the image-plane separation of the test and reference beams. Three methods for the removal of this error are proposed.

For several reasons, high-accuracy implementations of the EUV PDI and PS/PDI do not utilize *re-imaging optics* to image light from the exit pupil onto the detector plane. Such optics are common in most conventional interferometer designs. The primary reasons for their absence is the unavailability of optical elements of suitable quality and the fact that low-NA EUV measurements suffer only localized effects from diffraction. An important geometrical effect related to the absence of re-imaging optics causes a third-order systematic error to be introduced. Experimental observation of this effect has been used as a verification of system sensitivity (Section 8.9). The magnitude of this effect depends linearly on the image-plane separation of the test and reference beams, and thus affects both the PDI and the PS/PDI configurations.

Essentially, the test and reference beams are two diverging spherical beams with a lateral displacement of their centers-of-curvature. As they propagate toward the detector plane, the relative path-length difference generates the interference pattern. Neglecting aberrations in the optical system, the pattern consists primarily of parallel, uniformly-spaced, straight fringes; but consideration of the path-length difference including terms out to third order, reveals a systematic *coma* of magnitude comparable with the sensitivity of the EUV PS/PDI interferometer.

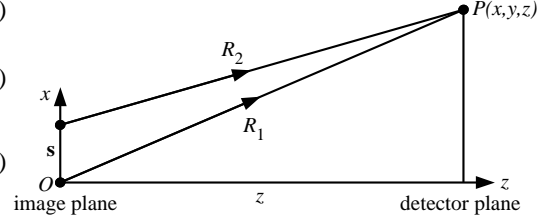
In Fig. 10, light from the two beams reaches a common point at the detector. Without loss of gener-

ality, we take the displacement vector  $\mathbf{s}$  to lie along the  $x$ -axis. Setting the origin on one of the rays, the path lengths are

$$R_1 = \sqrt{x^2 + y^2 + z^2}, \quad (44)$$

$$R_2 = \sqrt{(x-s)^2 + y^2 + z^2}, \quad (45)$$

$$\Delta R = z \sqrt{1 + \left(\frac{x}{z}\right)^2 + \left(\frac{y}{z}\right)^2} - z \sqrt{1 + \left(\frac{x-s}{z}\right)^2 + \left(\frac{y}{z}\right)^2}. \quad (46)$$



**Figure 10.** The description of systematic errors begins with the path length difference of the test and reference beam centers to a point on the detector. The figure shows the beam separation  $s$  and the image-to-detector-plane distance  $z$ .

At this point it is convenient to define the following dimensionless quantities

$$\begin{cases} u \equiv x/z \\ v \equiv y/z \\ \delta \equiv s/z \ll \max\{u\}, \max\{v\} \end{cases}. \quad (47)$$

Now re-write  $\Delta R$  in term of the dimensionless variables.

$$\Delta R = z \sqrt{1 + u^2 + v^2} - z \sqrt{1 + (u - \delta)^2 + v^2}. \quad (48)$$

Using the binomial expansion, expand the square-roots keeping terms up to first-order in  $\delta$  only. Many terms in the expansion cancel leaving

$$\Delta R = z \delta u \left[ 1 - \frac{1}{2}(u^2 + v^2) + \frac{3}{8}(u^2 + v^2)^2 - \dots \right] = \frac{z \delta u}{\sqrt{1 - (u^2 + v^2)}}. \quad (49)$$

To express this as a wavefront aberration observable in the data, it is convenient to use the normalized cylindrical *Detector Coordinate System* (see Section 5.2), normal to the  $z$ -axis.

$$\begin{cases} t \equiv \tan \alpha, \text{ where } \text{NA} \equiv \sin \alpha \\ \rho \equiv \frac{1}{t} \sqrt{u^2 + v^2} \in [0, 1], \text{ within the system NA} \\ \phi \equiv \tan^{-1}(v/u), \text{ the azimuthal angle} \end{cases}. \quad (50)$$

The path-length difference of interest may now be written as the product of radial and angular terms,

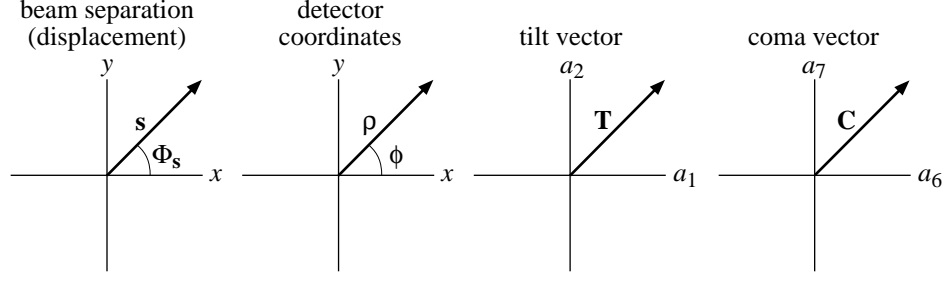
$$\Delta R = z \delta t \left[ 1 - \frac{1}{2} t^2 \rho^2 + \frac{3}{8} t^4 \rho^4 - \dots \right] \rho \cos \phi = \frac{z \delta t}{\sqrt{1 - t^2 \rho^2}} \rho \cos \phi. \quad (51)$$

Generalizing the direction  $\mathbf{s}$  as  $\phi_s$ , and replacing  $\delta$  with  $s/z$ ,  $\Delta R$  becomes

$$\Delta R = z \delta t \left( 1 - \frac{1}{2} t^2 \rho^2 + \frac{3}{8} t^4 \rho^4 - \dots \right) \rho \cos(\phi - \phi_s) = \frac{z \delta t}{\sqrt{1 - t^2 \rho^2}} \rho \cos(\phi - \phi_s), \quad (52)$$

$$\Delta R = s t \left( 1 - \frac{1}{2} t^2 \rho^2 + \frac{3}{8} t^4 \rho^4 - \dots \right) \rho \cos(\phi - \phi_s) = \frac{s t}{\sqrt{1 - t^2 \rho^2}} \rho \cos(\phi - \phi_s). \quad (53)$$





**Figure 11.** The description of several experimental quantities is facilitated by representation in pairs of coordinates. The test and reference beam separation in the image plane may be represented by a single vector  $\mathbf{s}$ . The tilt and coma components of the path length difference have a convenient representation in a single  $\mathbf{T}$  or  $\mathbf{C}$  vector defined from the Zernike polynomial coefficients  $(a_1, a_2)$  and  $(a_6, a_7)$  respectively. Since they share the same radial dependence, and differ only in the  $\cos\theta$  or  $\sin\theta$  angular dependence, these vector representations simplify many aspects of the analysis.

It will be useful to separate the angular dependence of Eq. (53) into cosine and sine components as follows.

$$\Delta R = st \left( 1 - \frac{1}{2} t^2 \rho^2 + \frac{3}{8} t^4 \rho^4 - \dots \right) \rho (\cos \phi_s \cos \phi + \sin \phi_s \sin \phi). \quad (54)$$

The first term in the expansion is the tilt that defines the fundamental fringe pattern. The negative sign of the third-order term in Eq. (54) shows that the effect of the geometric coma is a *reduction* of the fringe period at the edges of the measurement. (This also may be understood that from the perspective that at the edge of the field, a small change in angle results in a larger change in position on the detector than at the center.) The higher-order effects are always aligned parallel with the *tilt* term (also the beam separation), so there is no induced curvature of the fringes.

For a given optical system, the magnitude of this effect depends primarily on the image-plane separation of the test and reference beams. The bandwidth discussion of Section 5.4 showed that in the configuration where the beam from the first-diffracted order is used as the test beam, different wavelength components are brought to different image-plane separations. From the combination of these two effects, it is clear that attention to the chromatic dependence of the systematic coma may be necessary in some cases.

### 5.5.1 Representation of Zernike Pairs in Vector Notation

Further simplification of the path-length difference expansion can be made by introducing a vector notation for the pairs of Zernike terms that naturally separate into  $x$ - and  $y$ -oriented components. The definitions of the Zernike polynomials may be found in Chapter 14, and the Zernike coefficient-pair vector notation is discussed in Section 14.3.1. Here, the relevant terms are only the *tilt* and *coma* components.

Any wavefront aberration on an unobstructed circular aperture may be described by a series of Zernike polynomials, with coefficients  $\{a_i\}$ .

$$W(\rho, \phi) = \sum a_i Z_i(\rho, \phi). \quad (55)$$

The tilt and coma components are defined specifically as

$$\text{Tilt: } Z_1 \equiv \rho \cos \phi, \quad Z_2 \equiv \rho \sin \phi, \quad (56a)$$

$$\text{Coma: } Z_6 \equiv (3\rho^3 - 2\rho) \cos \phi, \quad Z_7 \equiv (3\rho^3 - 2\rho) \sin \phi, \quad (56b)$$

$\rho^3 \cos \phi$  and  $\rho^3 \sin \phi$  do not appear independently within the Zernike polynomials. A linear combination of tilt and coma is required to represent these terms.

$$\rho^3 \cos \phi = \frac{1}{3} Z_6 + \frac{2}{3} Z_1, \quad \text{and} \quad \rho^3 \sin \phi = \frac{1}{3} Z_7 + \frac{2}{3} Z_2. \quad (57)$$

For a simplified vector notation, define a position vector  $\mathbf{p}$

$$\mathbf{p} \equiv (\rho \cos \phi, \rho \sin \phi), \quad (58)$$

and two more vectors representing the tilt and coma coefficients of a Zernike Polynomials series.

$$\begin{cases} \mathbf{T} \equiv (a_1, a_2), & \text{“tilt vector”} \\ \mathbf{C} \equiv (a_6, a_7), & \text{“coma vector”} \end{cases} \quad (59)$$

These vectors are shown in Fig. 11.

Now, keeping only terms up to third-order, the path-length difference in Eq. (54) may be re-written. (The inclusion of higher-order terms, necessary only when NA or  $\mathbf{s}$  is large, is straightforward.)

$$\Delta R = t \left( 1 - \frac{1}{2} t^2 \rho^2 \right) \mathbf{p} \cdot \mathbf{s} = t (\mathbf{p} \cdot \mathbf{s}) - \frac{1}{2} t^3 (\rho^2 \mathbf{p} \cdot \mathbf{s}) = t \left( 1 - \frac{1}{3} t^2 \right) \mathbf{s} \cdot (Z_1, Z_2) - \frac{1}{6} t^3 \mathbf{s} \cdot (Z_6, Z_7). \quad (60)$$

$$\text{Hence,} \quad \mathbf{T} = t \left( 1 - \frac{1}{3} t^2 \right) \mathbf{s}, \quad \text{and} \quad \mathbf{C} = -\frac{1}{6} t^3 \mathbf{s}. \quad (61a) \text{ and } (61b)$$

Notice that  $\mathbf{s} \parallel \mathbf{T} \parallel \mathbf{C}$ . Finally, the path-length difference may be written as the sum of tilt and coma components.

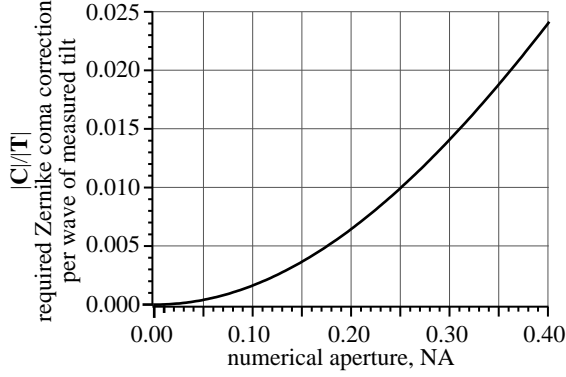
$$\Delta R = \mathbf{T} \cdot (Z_1, Z_2) + \mathbf{C} \cdot (Z_6, Z_7) = t \left( 1 - \frac{1}{3} t^2 \right) \mathbf{s} \cdot (Z_1, Z_2) - \frac{1}{6} t^3 \mathbf{s} \cdot (Z_6, Z_7). \quad (62)$$

### 5.5.2 Isolating and Removing the Geometric Coma Effect

Accurate PDI or PS/PDI wavefront measurement in the absence of re-imaging optics requires that the systematic error from the geometrical coma be identified and subtracted from the data. There are several means available for determining the magnitude of this effect. Two methods are described here.

The magnitude of the geometric coma depends very sensitively on the NA of measurement ( $\text{NA}^3$  dependence). When the data is analyzed, this NA is not strictly the NA on the measurement-side of the optical system. Typically a sub-region of the available data is selected: the relevant NA of interest here is the NA *defined* by the selected sub-region and the cone of rays that created it. In practice it may therefore be difficult to precisely know the *measurement NA*.

Method 1 outlines a procedure to follow when the measurement NA is well known. In Method 2, two separate measurements with different fringe rotations and/or densities are combined, and *a priori* knowledge of the measurement NA is *not* required. In both cases, the goal is to determine the change



**Figure 12.** The image plane separation of the test and reference beams introduces a systematic geometric coma error. This graph shows the magnitude of the required Zernike coefficient coma correction per wave of measured tilt as a function of NA. At 0.08 NA, the correction is 0.0011 waves of coma is required for each wave of measured tilt. At 0.1 NA, the correction is 0.0017 waves per wave of tilt.

required to remove the geometric coma from the Zernike polynomial series, or, equivalently, from the measured wavefront itself.

**5.5.2.1 Method 1: Removing the Geometric Coma with Known Measurement NA.** If the measurement NA is precisely known, then the removal of the geometric coma systematic error is straightforward. Analysis proceeds from the path-length difference of Eq. (62). The tilt and coma vectors are parallel and have a fixed relationship based on  $t$ , the tangent of the NA angle.

$$\mathbf{C} = -\frac{t^2 \mathbf{T}}{6(1 - \frac{1}{3}t^2)} = -\frac{t^2 \mathbf{T}}{6 - 2t^2} \approx -\frac{1}{6}t^2 \mathbf{T}. \quad (63a)$$

$$\mathbf{C} = -\frac{NA^2 \mathbf{T}}{6 - 8NA^2} \approx -\frac{1}{6}NA^2 \mathbf{T}, \quad (63b)$$

In terms of the NA,

The approximation holds for small NA.

In the presence of wavefront aberrations, the measured coma  $\mathbf{C}_o$  may take any arbitrary value. From this coma, the geometric coma error  $\mathbf{C}$  must be subtracted to yield  $\mathbf{C}'$  the *actual* coma. Using the measured tilt and the known NA, the geometric coma subtraction is as follows:

$$\mathbf{C}' = \mathbf{C}_o - \mathbf{C} = \mathbf{C}_o - \frac{t^2 \mathbf{T}}{6 - 2t^2}. \quad (64)$$

Figure 12 shows the significance of this correction by plotting the amount of coma correction required (in waves) per wave of measured tilt. If the system has 40 fringes, multiply the ordinate by 40 to find the magnitude of required coma correction in waves.

Section 5.9 describes a coma that comes from the planar grating diffracting spherical beam. If this effect is present in the test wavefront, then Eqns. (63) and (64) may need modification to account for this effect. Like the geometric coma, the grating coma is also proportional to the tilt, so the modification is not complicated.

**5.5.2.2 Method 2: Removing Geometric Coma Using the Difference Wavefront.** Performing two separate experiments using gratings aligned at different angles or with different pitch, enables a combina-

tion of measurements that can be used to identify and remove the geometric coma. (The image-plane reference pinhole(s) and window must be designed to accommodate this.) This analysis method utilizes a Zernike polynomial fit to the *difference wavefront*, representing the difference between two separate measurements. Equivalently, since the fitting polynomials are orthogonal, the first set of coefficients may be simply subtracted from the second set to provide the fitting coefficients of the difference wavefront.

Consider separate measurements using two gratings inserted into the same beam position, normal to the central ray, but with the rulings oriented along different directions. The image-plane beam separations will be  $\mathbf{s}_1$  and  $\mathbf{s}_2$ , not necessarily equal in magnitude. Assume that the optical system under test has an arbitrary wavefront aberration  $W(\rho, \phi)$ . To reduce measurement uncertainties, the two wavefronts used here may themselves be composite wavefronts formed from multiple series of similar measurements.

If for both measurements the test beam passes through the same image-plane point, then the contribution of tilt and geometric coma to the two observed path length differences may be written according to Eq. (59). Including arbitrary wavefront aberrations  $W$ ,

$$\Delta R_1 = W + t\left(1 - \frac{1}{3}t^2\right)\mathbf{s}_1 \cdot (Z_1, Z_2) - \frac{1}{6}t^3\mathbf{s}_1 \cdot (Z_6, Z_7) = \mathbf{T}_1 \cdot (Z_1, Z_2) + \mathbf{C}_1 \cdot (Z_6, Z_7), \quad (65a)$$

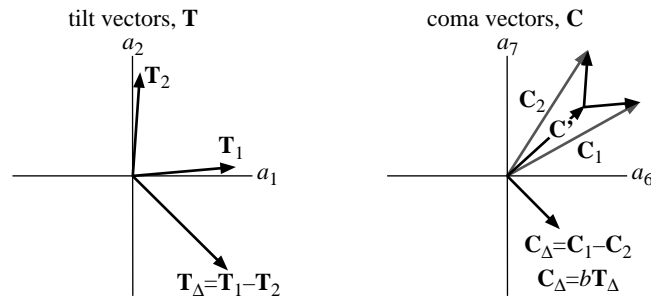
$$\Delta R_2 = W + t\left(1 - \frac{1}{3}t^2\right)\mathbf{s}_2 \cdot (Z_1, Z_2) - \frac{1}{6}t^3\mathbf{s}_2 \cdot (Z_6, Z_7) = \mathbf{T}_2 \cdot (Z_1, Z_2) + \mathbf{C}_2 \cdot (Z_6, Z_7). \quad (65b)$$

Taking the difference,

$$\Delta R_1 - \Delta R_2 = (\mathbf{T}_1 - \mathbf{T}_2) \cdot (Z_1, Z_2) + (\mathbf{C}_1 - \mathbf{C}_2) \cdot (Z_6, Z_7) = \mathbf{T}_\Delta \cdot (Z_1, Z_2) + \mathbf{C}_\Delta \cdot (Z_6, Z_7), \quad (66)$$

where  $(\Delta R_1 - \Delta R_2)$  is the measured difference wavefront, and  $\mathbf{T}_\Delta$  and  $\mathbf{C}_\Delta$  are the *measured* Zernike coefficients that describe it.

$$\begin{cases} \mathbf{T}_\Delta \equiv (\mathbf{T}_1 - \mathbf{T}_2) = (a_1, a_2) \\ \mathbf{C}_\Delta \equiv (\mathbf{C}_1 - \mathbf{C}_2) = (a_6, a_7) \end{cases} \quad (67)$$



**Figure 13.** When two or more measurements are made at different orientations of the spatial carrier frequency of the fringe pattern, the systematic coma may be isolated and removed. This is facilitated by the definition of two *difference vectors*  $\mathbf{T}_\Delta$  and  $\mathbf{C}_\Delta$ , as shown. The systematic coma components must be parallel to the tilt in the individual measurements. Experimentally the measured comas  $\mathbf{C}_1$  and  $\mathbf{C}_2$  come from the inherent coma  $\mathbf{C}'$  plus the systematic coma, oriented parallel to the tilt in each measurement. By using the difference coma  $\mathbf{C}_\Delta$ , the inherent coma is separated from the systematic coma components. The proportionality constant  $b$  between  $\mathbf{C}_\Delta$  and  $\mathbf{T}_\Delta$  is easily found using a least-squares technique.

The contributions of the wavefront we are trying to measure  $W$  are removed by the subtraction.

Figure 13 shows the tilt and coma vectors of two separate measurements, and the difference vectors described by Eq (67). The measured coma vectors  $\mathbf{C}_1$  and  $\mathbf{C}_2$  both contain the inherent coma  $\mathbf{C}'$ , which is removed by subtraction.

Even though the NA (and thus  $t$ ) is not known, we may utilize the fact that the *ratio* of  $\mathbf{T}_1$  to  $\mathbf{C}_1$  and of  $\mathbf{T}_2$  to  $\mathbf{C}_2$  is fixed, and solve for the proportionality constant  $b$  that provides the best fit. Using the method of least-squares, the criterion for the best fit to the data is to find the minimum of the error function  $E^2(b)$ .

$$E^2(b) = \|\mathbf{C}_\Delta - b\mathbf{T}_\Delta\|^2 = \|\mathbf{C}_\Delta\|^2 + b^2\|\mathbf{T}_\Delta\|^2 - 2b(\mathbf{T}_\Delta \cdot \mathbf{C}_\Delta). \quad (68)$$

The minimum occurs where the derivative with respect to  $b$  is zero.

$$\frac{dE^2(b)}{db} = 0 = 2b\|\mathbf{T}_\Delta\|^2 - 2(\mathbf{T}_\Delta \cdot \mathbf{C}_\Delta). \quad (69)$$

$$b = \frac{\mathbf{T}_\Delta \cdot \mathbf{C}_\Delta}{\|\mathbf{T}_\Delta\|^2} = \frac{a_1 a_6 + a_2 a_7}{a_1^2 + a_2^2}. \quad (70)$$

Solving for  $b$ ,

By the known relationship of tilt and coma in Eq. (63), we can solve for  $t$ , and thus NA.

$$b = -\frac{t^2}{6 - 2t^2}, \text{ and } t = \sqrt{\frac{6b}{2b - 1}}. \quad (71) \text{ and } (72)$$

$$\text{NA} = \sin(\tan^{-1} t) = \frac{t}{\sqrt{1 + t^2}}. \quad (73)$$

From here, the procedure for removing the coma follows Method 1. From the two measured comas  $\mathbf{C}_{1,2}$  and the measured tilts  $\mathbf{T}_{1,2}$  the geometric coma is subtracted. Separately, for each measurement (1 and 2), we find the underlying coma  $\mathbf{C}'$ .

$$\mathbf{C}' = \mathbf{C}_{1,2} - \mathbf{C} = \mathbf{C}_{1,2} - b\mathbf{T}_{1,2} = \mathbf{C}_{1,2} - \left( \frac{a_1 a_6 + a_2 a_7}{a_1^2 + a_2^2} \right) \mathbf{T}_{1,2}. \quad (74)$$

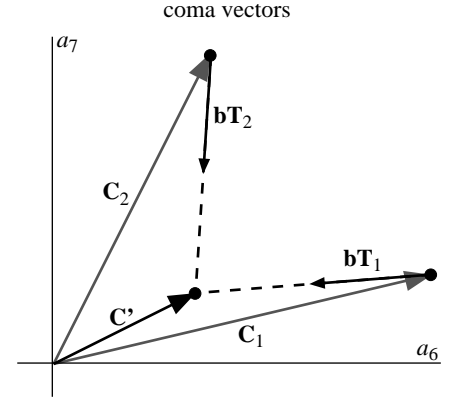
The measured wavefront, after the removal of the systematic coma, is found from the *average* or another suitable combination of the two sets of measurements.

As stated earlier, if the so-called *grating coma* systematic error is present in the test wavefront, then the above discussion requires some simple modifications. The addition of the grating coma only affects the proportionality constant between the tilt and coma terms. Since for Method 2, the measurement NA is a parameter of the fit, and the proportionality constant is unknown, no modification is required to determine the inherent coma. However, Equations (71) through (73) which relate the fitting parameter  $b$  to the measurement NA, do require modification.

### 5.5.2.3 Alternate description of Method 2: Removing Geometric Coma

There is an alternate geometric description of the coma subtraction in Method 2 that does not use the difference wavefront, but yields the same solution. We utilize the two separate measurements, and the fact that the tilt vectors in each must be proportional to the geometric coma, *with the same proportionality constant*.

Figure 14 shows a graphical representation of this method. Using the coma terms from two measurements performed at different beam displacements  $C_1$  and  $C_2$ , vectors proportional to the tilts  $T_1$  and  $T_2$  are subtracted to reach the best agreement. The distance between the two points is minimized (they may not match exactly) at the location of the inherent coma  $C'$  we are trying to find. Following the least-squares method we define an error function  $E^2(b)$  that here represents a distance in the coefficient *vector* space shown in Fig. 14.



**Figure 14.** A geometrically alternate, yet mathematically identical description of the systematic coma removal. Here the measured coma,  $C_1$  and  $C_2$ , contain both the inherent coma  $C'$  and the systematic coma components that are parallel to the measured tilts  $T_1$  and  $T_2$ . The constant of proportionality  $b$  depends only on the NA and is the same for both. Finding the  $b$  that provides the best agreement between the inherent coma of the two separate measurements yields  $C'$ .

$$E^2(b) = \|(C_1 - bT_1) - (C_2 - bT_2)\|^2 = \|C_\Delta\|^2 + b^2\|T_\Delta\|^2 - 2b(T_\Delta \cdot C_\Delta). \quad (75)$$

This expression is identical to Equation (68), and thus its solution will be the same.

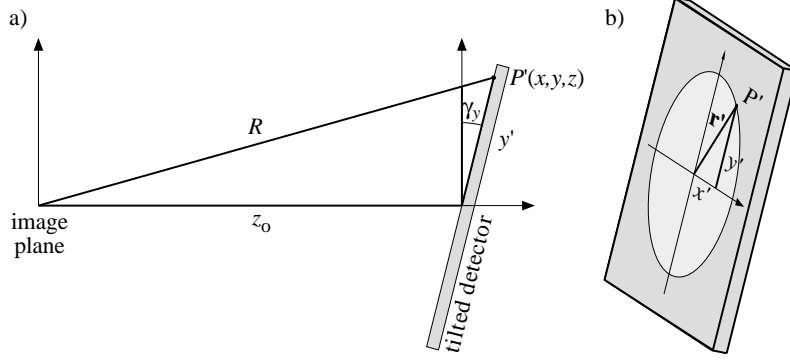
### 5.5 Summary

**Measured Geometric Coma.**  $|C| \approx 1/6 \text{ NA}^2 |T| \Rightarrow$  At 0.08 NA,  $|C|/|T| = 1/6 \cdot 0.08^2 = 0.0011$  waves per wave of tilt =  $5.5 \times 10^{-4}$  waves per fringe.  $|C| = 0.37 \text{ nm}$  @ 50 fringes. At 0.1 NA,  $|C|/|T| = 1/6 \cdot 0.1^2 = 0.0017$  waves per wave of tilt =  $8.3 \times 10^{-4}$  waves per fringe.  $|C| = 0.56 \text{ nm}$  @ 50 fringes.

## 5.6 SYSTEMATIC ERROR FROM DETECTOR MISALIGNMENT

It is reasonable to assume that the planar detector used in PS/PDI interferometry is not perfectly aligned, with its surface-normal parallel to the central ray of the optical system. Such misalignment, represented as a small inclination of the detector plane, introduces a systematic astigmatic error. The magnitude of this error depends on the beam separation and may be comparable to the target accuracy. The sensitivity of a given configuration to detector misalignment is presented at the conclusion of this section.

Following Section 6.5 on the geometrical coma systematic error, the effect of the detector misalignment on the observed interference pattern may be derived in terms of its effect on the path-length difference between the test and reference beams, observed in a coordinate system appropriate for the detector.



**Figure 15.** Definition of the coordinate systems used in the description of the detector misalignment systematic error.

Begin with Equation (51) for the path length difference, using  $z_o$  as the distance from the image-plane to the detector along the  $z$ -axis, as shown in Fig. 15. In terms of the dimensionless coordinates, the path length difference  $\Delta R_o$  is written in the case of perfect alignment

$$\Delta R_o = z_o \delta t \left( 1 - \frac{1}{2} t^2 \rho^2 \right) \rho \cos \phi, \quad (76)$$

where, without loss of generality, the  $x$ -axis is defined along the displacement  $\mathbf{s}$  of the test and reference beams. As before,  $t$  is the tangent of the NA angle,  $\delta$  is the dimensionless angle related to the beam separation, and  $\rho$  is a dimensionless radial coordinate in the detector system. Maintaining the cylindrical coordinate system, and reintroducing  $r = \rho z_o$  as the *regular, Laboratory* radial coordinate,

$$\Delta R_o = \frac{s}{z_o} \left( 1 - \frac{r^2}{2z_o^2} \right) r \cos \phi. \quad (77)$$

Figure 15 shows how the coordinate systems are defined.  $r'$  represents the radial coordinate in the detector plane, while  $r$  is the real-space radial coordinate. With a non-zero detector tilt angle  $\gamma$ , there are small changes in  $z$ ,  $x$ , and  $y$  across the detector. Define the vectors  $\mathbf{r}' \equiv (x', y')$  in the detector plane, and  $\mathbf{r} \equiv (x, y)$  in the Laboratory System, and, as before, the polar *angular vector*  $\boldsymbol{\gamma} \equiv (\gamma_x, \gamma_y)$ . Based on the tilt angle  $\boldsymbol{\gamma}$ , misalignment of the detector introduces a first-order change in  $z$  and a second-order change in the lateral coordinates. Assuming small misalignments, only terms up to first order in  $\boldsymbol{\gamma}$  will be kept in the following discussion.

$$z \approx z_o + \gamma_x x' + \gamma_y y' = z_o + \boldsymbol{\gamma} \cdot \mathbf{r}' = z_o \left( 1 + \frac{\boldsymbol{\gamma} \cdot \mathbf{r}'}{z_o} \right). \quad (78)$$

The new path length difference becomes

$$\Delta R = \frac{s}{z} \left( 1 - \frac{r^2}{2z^2} \right) r \cos \phi. \quad (79)$$

Using the first-order expansion of  $z$  in  $\boldsymbol{\gamma}$  from Eq. (78), Eq. (79) gives

$$\Delta R \approx \frac{1}{z_o} \left( 1 - \frac{\boldsymbol{\gamma} \cdot \mathbf{r}'}{z_o} \right) \left[ 1 - \frac{r^2}{2z_o^2} \left( 1 - \frac{2\boldsymbol{\gamma} \cdot \mathbf{r}'}{z_o} \right) \right] sr \cos \phi. \quad (80)$$

Keeping only the most significant terms, Eq. (80) becomes

$$\Delta R \approx \Delta R_o - \frac{\gamma \cdot \mathbf{r}'}{z_o^2} s r \cos \phi . \quad (81)$$

Putting this back in terms of the dimensionless coordinates ( $\rho, \phi, z$ ),

$$\Delta R \approx \Delta R_o - t^2 s (\gamma \cdot \rho) \rho \cos \phi . \quad (82)$$

To simplify this expression, redefine how the detector tilt is described: the detector is inclined by an angle  $\gamma$ , in the azimuthal direction  $\phi_\gamma$ . Then

$$\gamma \cdot \rho = \gamma \rho \cos(\phi - \phi_\gamma) , \quad (83)$$

and Eq. (82) becomes

$$\Delta R \approx \Delta R_o - s \gamma t^2 \rho^2 \cos(\phi - \phi_\gamma) \cos \phi = \Delta R_o - \frac{1}{2} s \gamma t^2 \rho^2 \left[ \cos(2\phi - \phi_\gamma) + \cos \phi_\gamma \right] . \quad (84)$$

This has the effect of adding a small astigmatism to the measurements. Isolating the difference from  $\Delta R_o$  leaves

$$\Delta R' \approx -\frac{1}{2} s \gamma t^2 \rho^2 \left[ \cos(2\phi - \phi_\gamma) + \cos \phi_\gamma \right] . \quad (85)$$

The presence of the constant  $\cos \phi_\gamma$  term adds defocus and makes the magnitude of this effect different when the tilt direction is parallel or perpendicular to the beam separation direction. A detector tilt in the  $s$ -direction ( $x \parallel s$ ) produces a “cylindrical” path length difference of

$$x\text{-tilt:} \quad \Delta R'_x \approx -\frac{1}{2} s \gamma t^2 \rho^2 (\cos 2\phi + 1) . \quad (86)$$

For a tilt in the  $y$ -direction ( $y \perp s$ ), the path-length change is astigmatic.

$$y\text{-tilt:} \quad \Delta R'_y \approx -\frac{1}{2} s \gamma t^2 \rho^2 \sin 2\phi . \quad (87)$$

The term in Eq. (86) behaves as a small defocus, arising from the fact that one of the beams is off-axis.

### 5.6.1 Numbers

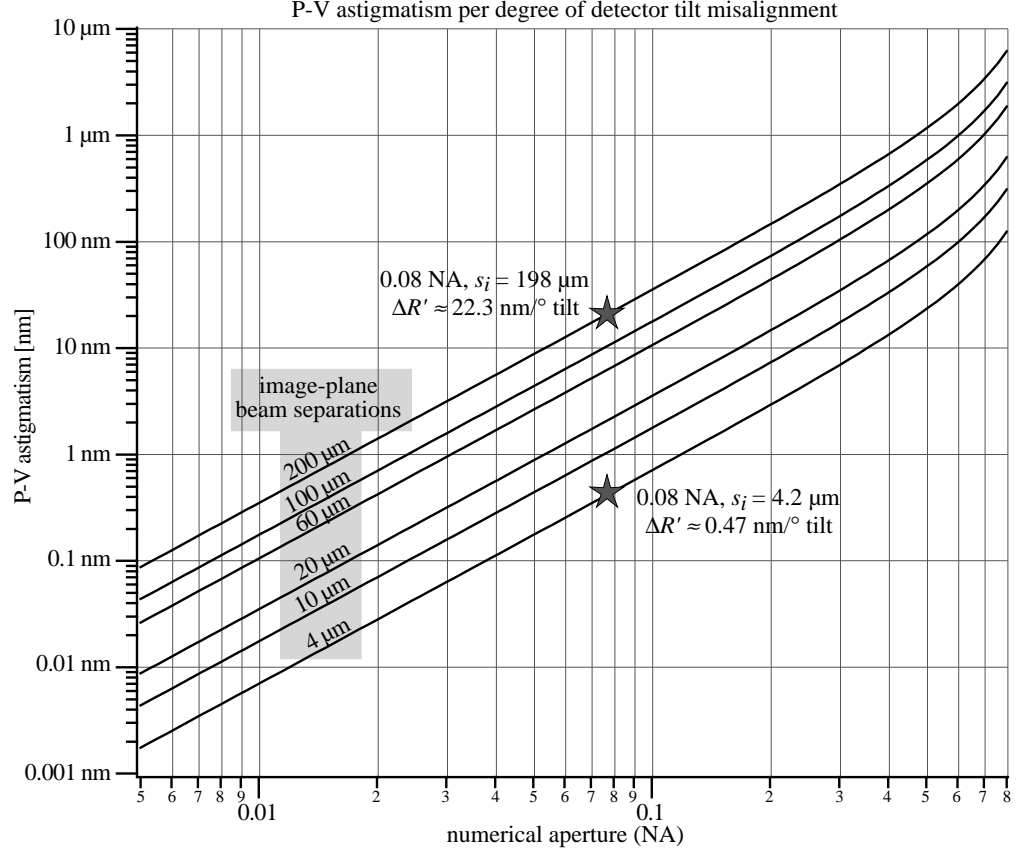
The peak-to-valley magnitude of the astigmatism described by Eq. (87) is

$$\text{peak-to-valley:} \quad |\Delta R'_y| \approx s \gamma t^2 \approx s \gamma N A^2 . \quad (88)$$

The approximation holds for small NA. Equation (88) is plotted in Fig. 16 as a function of NA for beam separations in the range relevant to EUV, visible, and 193 nm system measurements. For convenience, we can re-write Eq. (88) putting  $\gamma$  in degrees rather than radians. The peak-to-valley astigmatism magnitude per degree of detector tilt is

$$\text{peak-to-valley:} \quad |\Delta R'_y| \approx 0.0175 s N A^2 \approx 1.1 \times 10^{-4} s @ N A = 0.08 . \quad (89)$$





**Figure 16.** Detector misalignment (tilt) introduces a path-length difference between the test and reference beams causing a systematic astigmatic error dependent on the beam separation and the measurement NA. The gray stars indicate relevant values for EUV and visible-light PS/PDI interferometry at 0.08 NA.

### 5.6 Summary

**Detector Misalignment.** P-V astigmatism  $A = s_i \gamma NA^2 \Rightarrow \sim 0.47 \text{ nm/}^\circ \text{ tilt}$ . Also,  $A/\gamma N_{\text{fringe}} = \lambda NA/2 = 0.54 \text{ nm/}^\circ \text{ tilt/fringe}$ . The measured Zernike coefficient of astigmatism is half of this, or  $0.27 \text{ nm/}^\circ \text{ tilt/fringe}$ .

### 5.7 GRATING PLACEMENT CONSIDERATION: SHEAR

As discussed in Section 4.4, the PS/PDI shares many similarities to a conventional lateral shearing interferometer (LSI) in that both systems introduce a relative beam *shear*, or displacement, to generate the interference pattern. In principle, the various configurations of the LSI interfere the test beam with a sheared copy (or copies) of itself. The PS/PDI, on the other hand, produces a spherical reference wavefront by spatially filtering one copy of the test beam in the image-plane where the beams are separated. When the measurement involves spherically diverging beams and no re-imaging optics, in both configurations the central rays of the two beams are directed at slightly different angles. The beam shear in the PS/PDI is determined by the grating pitch and the illumination wavelength. A comparison of the importance of shear

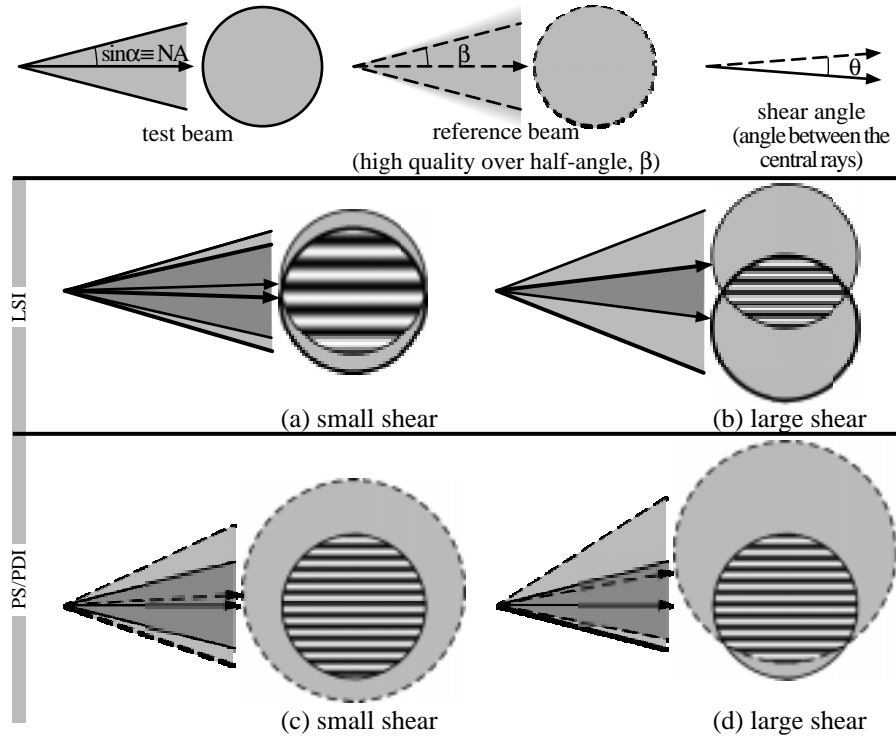
in the two interferometers is shown in Fig. 17.

By rule, the PS/PDI reference pinhole should be chosen small enough that the reference wavefront *significantly overfills* the measurement NA. However, since pinhole-diffracted reference wavefronts are typically of suitably high quality only within a cone of finite angle, selection of the grating position and pitch should be made with attention to the beam shear angle. For a given optical system and wavelength, the number of fringes in the interferogram depends only on the image-plane separation of the test and reference beams (Section 5.3). There are, however, infinite combinations of grating pitch and position that yield the same separation.

Here, the discussion is limited to the PS/PDI configuration with the grating placed between the test optic and the image-plane. Similar analysis for other PS/PDI configurations follow a nearly identical form: where the grating is placed before the test optic, the shear angle is scaled by the system magnification.

From the grating equation, the shear angle  $\theta$  is equal to  $\lambda/d$ . For a grating of pitch  $d$  and distance from the image-plane  $z$  the image- or object-plane separation of beam is

$$\theta z \approx \frac{\lambda z}{d} \approx s. \quad (90)$$



**Figure 17.** The importance of beam shear considerations is shown in this figure. Shear is here defined as the angle between the central rays of the test and reference beams. The test beam is represented with solid lines, and the reference beam with dashed lines, as shown in the top row. In the LSI (a) and (b), the test beam interferes with a sheared copy of itself, and fringes are produced in the overlap region. When the shear is large (b), only a fraction of the available area is investigated. In the PS/PDI the spatially-filtered reference beam should have an NA larger than the test optic (c). However, if the shear angle is large (d), then to guarantee that the reference beam will overlap the test beam over the measurement NA, the requirement on the diffraction angle of the reference beam becomes more severe.

Assume for a given application that the reference wavefront is of arbitrarily high quality only over a cone defined from the central ray, out to a half-angle  $\beta$ . Clearly, a minimum requirement for measurement is that  $\beta > \alpha$ , the maximum half-angle within the NA of measurement. When the shear angle  $\theta$  is significant relative to  $\alpha$ , and the test and reference beams are displaced, the new requirement on  $\beta$  becomes

$$\beta > \alpha + \theta . \quad (91)$$

Producing a high-quality reference wavefront is a matter of foremost importance and a significant challenge to point-diffraction interferometry. Any method of relaxing the requirements on the magnitude of  $\beta$  gives more freedom to other experimental parameters. One direct means of reducing  $\beta$  is to keep the shear angle  $\theta$  as small as possible. For a given image- or object-plane beam separation  $s$ ,  $\theta$  may be reduced by moving the grating away from the image-plane (or away from the object plane in other PS/PDI configurations). Choosing the optimum grating position requires balances the often opposing concerns of the grating's pitch and the illuminated area. Fabrication issues may constrain the maximum size of the grating, but gratings of larger pitch (coarser) may often be made to higher quality.

### 5.7 Summary

**Shear Angle.**  $\theta = \lambda/d =$  shear angle.  $d =$  grating pitch.  $\beta =$  half-angle over which reference wavefront is of arbitrarily high quality.  $\alpha =$  NA. Minimum requirement:  $\beta > \theta + \alpha$ .

## 5.8 GRATING FABRICATION ERRORS

Aberrations and local imperfections in the grating-beamsplitter can contribute directly measurement errors. This section describes the most significant grating error contributions, and recommends various methods of overcoming them. The most important recommendation is that when the quality of the grating cannot be guaranteed to beyond the level of measurement accuracy desired, then one of the first-diffracted order beams should be filtered to become the reference beam.

It is helpful to view the grating, which serves a dual role as beamsplitter and phase-shifting element, as a binary transmission hologram approximating the coherent interference of multiple plane waves separated by small angles. Imperfections in the grating pattern can be described by aberrations in the interfering beams. The *inversion* of this description (by Babinet's Principle) is a single illuminating beam diffracted by the imperfect grating into multiple, coherent, aberrated beams.

For the following discussion it is useful to treat different types of grating imperfections separately. Figure 18(b) shows several types of grating defects. Pattern placement errors, in which the unbroken opaque lines are not accurately drawn, are referred to as *grating aberrations*. The other kinds of defects in which the opaque lines are missing, transparent regions are obstructed, or the thickness of the supporting substrate

is non-uniform are addressed separately from the grating aberrations.

### 5.8.1 Grating Aberrations

It is important to note that the phase of the zeroth-order diffracted beam is *not affected* by the grating pattern itself. Light propagating into the zeroth-order adds *in-phase*, independent of the positions of the rulings. This is because the grating pattern introduces no path-length change into the various parts of the zeroth-order beam. On the other hand, the diffracted beams are *defined by* the grating positions: the wavefront phase of these diffracted beams is subject directly to the grating aberrations. The following discussion presents a simple analogy that is used to demonstrate this point.

In the absence of a grating, the superposition of multiple coherent beams would form a stationary intensity pattern in the grating plane. By Babinet's Principle (Babinet 1837), the single-beam illumination of a grating that approximates this same intensity pattern generates the diffraction of multiple beams similar to the former configuration.

To illustrate this point, consider a grating of pitch  $d$  with rulings aligned perpendicular to the  $x$ -axis. Define the spatial carrier frequency of the grating  $\kappa$

$$\kappa = \frac{2\pi}{d} \hat{\mathbf{x}}. \quad (92)$$

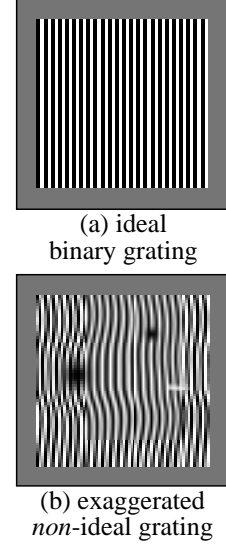
We may represent the grating transmission function  $T(\mathbf{r})$  as a *square wave* defined by an arbitrary, spatially varying grating phase  $\Phi(\mathbf{r})$ . Separating the grating phase into an aberration function  $\phi(\mathbf{r})$  and a carrier frequency, we have

$$\Phi(\mathbf{r}) = \phi(\mathbf{r}) + \kappa \cdot \mathbf{r} = \phi(\mathbf{r}) + \kappa x, \quad (93)$$

and

$$T(\mathbf{r}) = \begin{cases} 1, & \cos[\Phi(\mathbf{r})] \geq 0 \\ 0, & \cos[\Phi(\mathbf{r})] < 0 \end{cases}. \quad (94)$$

This description leads naturally to a representation of the grating as the interference of two beams. At this point, we neglect the spherically diverging or converging angle of incidence, and consider the beams as plane waves. (Section 5.9 addresses the systematic error issues related to the planar grating in a



**Figure 18.** Imperfections in the PS/PDI grating beamsplitter can introduce aberrations into the test and reference beams. (a) shows an ideal binary transmission grating of equal line and space ratio. The grating in (b) contains several aberrations types discussed in this section: low-spatial frequency pattern errors, dust or substrate errors, and pattern defects. These types of aberrations can affect the test and reference beam in different ways.

spherical beam.) In this description, assume unit intensity of the beams, neglect variation of the beam intensities, and assume that the test beam is unaberrated. The normalized intensity of the two interfering beams is given by

$$I(\mathbf{r}) = \frac{1}{4} \left| 1 + e^{i\Phi(\mathbf{r})} \right|^2 = \frac{1}{2} + \frac{1}{2} \cos[\Phi(\mathbf{r})]. \quad (95)$$

The square-wave grating transmission function of Eq. (94) is an approximation to this sinusoidal variation. Placing a simple threshold on Eq. (95) completes the analogy to Eq. (94), and justifies the approximation. Since  $\Phi$  was chosen arbitrarily, then for any grating phase function  $\Phi(\mathbf{r})$  the diffracted beams acquire a wavefront aberration  $\phi(\mathbf{r})$  and a direction determined by  $\mathbf{\kappa}$ .

Regarding the description of spherical beams, Eq. (95) may be generalized to allow both of the interfering beams to contain an additional phase term representing the path length difference between a spherical surface and the grating plane. This additional phase, defined as  $\Gamma(\mathbf{r})$ , appears in both beams, and thus the resultant intensity pattern is unaffected. Mathematically,

$$I'(\mathbf{r}) = \frac{1}{4} \left| e^{i\Gamma(\mathbf{r})} + e^{i\Phi(\mathbf{r}) + i\Gamma(\mathbf{r})} \right|^2 = \frac{1}{2} + \frac{1}{2} \cos[\Phi(\mathbf{r})] = I(\mathbf{r}). \quad (96)$$

The analogy may be extended to include the higher diffracted orders. Re-creation of a square-wave intensity profile in the grating plane requires an infinite series of interfering beams, each with the same phase aberration, but with a different propagation direction and intensity. These beam directions are given by positive and negative integer multiples of  $\mathbf{\kappa}$  (. . .  $-2\mathbf{\kappa}$ ,  $-\mathbf{\kappa}$ ,  $0$ ,  $\mathbf{\kappa}$ ,  $2\mathbf{\kappa}$ , . . .  $n\mathbf{\kappa}$ , . . .). The illumination of the square-wave grating with the single test beam generates this same series of diffracted orders. This series may be generalized as a Fourier cosine series.

$$T(\mathbf{r}) = \sum_n a_n \cos[\phi(\mathbf{r}) + n\mathbf{\kappa}x]. \quad (97)$$

### 5.8.2 Phase-Shifting

The origin of the phase-shifting properties of the grating is easily shown from the discussion of the previous section. Here, neglect aberrations and imperfections in the grating, and assume that the grating is defined by a carrier-frequency  $\mathbf{\kappa}$  and a square-wave transmission function. Once again we equate the coherently-combined intensity of a series of interfering beams with the grating transmission function. As above, taking  $\mathbf{\kappa}$  to be aligned with the  $x$ -axis, the translational invariance of the grating along the  $y$ -direction allows the substitution of  $x$  for  $\mathbf{r}$ .

$$I(x) = T(x) = \sum_n a_n \cos(n\mathbf{\kappa}x). \quad (98)$$

Physical translation of the grating in the  $x$ -direction, perpendicular to the grating rulings, may be expressed as

$$I(x - x_o) = T(x - x_o) = \sum_n a_n \cos[n\kappa(x - x_o)] = \sum_n a_n \cos(n\kappa x - \Delta_n) \quad , \quad (99)$$

where we defined the phase step as  $\Delta n \equiv n\kappa x_o$ . This very important result shows that a physical translation of the grating produces the same effect on the diffracted waves as a constant phase offset between the interfering beams in our model. Furthermore, between any two *adjacent* grating orders ( $\Delta n = \pm 1$ ), for a given grating translation the relative phase shift will be the same. That is, the expression

$$\Delta\phi(n, n-1; x_o) = n\kappa x_o - (n-1)\kappa x_o = \kappa x_o \quad (100)$$

is invariant in  $n$ .

Another important, albeit obvious, by-product of this discussion is that the grating translation required to produce a  $2\pi$ , single-cycle, phase shift between adjacent diffracted beams is simply  $d$ . (Recalling the definition of  $\kappa \equiv 2\pi/d$ , we can see from Eq. (100) that the translation  $x_o$  required to produce a  $2\pi$  phase change is in fact  $d$ .) Since the ideal grating in this treatment is periodic in  $x$  with period  $d$ , we should expect that translation by  $d$  returns the system to its original state.

### 5.8.3 Local Imperfections and Substrate Errors

Besides the low-spatial frequency pattern errors which introduce phase aberrations into the diffracted beams, there are other, higher-spatial frequency errors of concern. An opaque dust particle or a defect within the illuminated area of the grating may appear as a dim region, or a region of low (or zero) fringe contrast in the data. Since the plane of the grating is not typically imaged onto the detector plane, diffraction broadens the features of these high-spatial frequency aberrations. As the grating is translated over several fringe cycles, the motion of these aberrations will distinguish them from the stationary defects in the optics or elsewhere. By performing careful measurement, it may be possible to overcome localized grating defects by using other *clean* regions of the grating.

One form of grating fabrication error is perhaps the most troublesome. If the grating is patterned on a membrane or substrate, then substrate thickness variations can introduce phase errors that could be very difficult errors to overcome. In that case, the quality of the test beam is directly compromised. Once again, careful measurements performed using different regions of a large grating may reveal the presence of such systematic errors.

### 5.8.4 Recommendations

The above discussion leads to a recommendation that may appear counter-intuitive. It has been shown that aside-from local imperfections and substrate errors, grating pattern aberrations create phase errors only in the diffracted beams. By allowing the zeroth-order to become the test beam, and by spatially filtering one of the diffracted orders to become the reference beam, these grating aberrations may be over-

come. That is, where concerns about the grating quality exist, the first-order reference configuration is recommended over the zeroth-order reference configuration. This recommendation, however, runs contrary to the assertion that the zeroth-order reference configuration should be used to achieve high fringe-contrast.

### 5.8 Summary

**Grating Fabrication Errors.** Recommendation: use the *first-order reference* configuration whenever grating fabrication error magnitudes are unknown, or are known to be comparable with the desired accuracy.

## 5.9 GRATING COMA

Another potentially significant systematic error comes from the use of a planar grating beamsplitter in a spherically diverging or converging beam. Since the angles of incidence vary across the illuminated region of the grating, a small phase error is introduced into the diffracted beams. In a geometrical description, the grating pitch appears reduced to the off-axis rays perpendicular to the grating rulings. This leads to a variation in the diffraction angle within the cone of the beam.

The *grating coma* introduced here may be filtered, and therefore eliminated, when the interferometer is used in the first-order reference configuration — the grating aberrations are manifest only in the non-zero diffracted orders (Section 5.8). Therefore with appropriate filtering, the relevance of the grating coma may be limited to the zeroth-order reference configuration only.

The mathematical formulation presented here follows from the discussion of grating aberrations in Section 5.8. It is important here to consider the spherical divergence of the illuminating beam. As before, we create an analogy between the grating transmission, and the intensity pattern produced by a pair of coherently interfering beams in the grating plane. Starting with a single, illuminating beam, we solve for the phase aberrations of a *diffracted beam* required to produce the desired pattern. Limiting our discussion to the interference of only two beams simplifies the problem considerably. To that end, consider only the fundamental sinusoidal-transmission of an ideal grating of pitch  $d$ . The descriptions for *converging* and *diverging* beams, with a radius of curvature  $R$  are identical in form.

Consider the illuminating beam to be an ideal spherical wave diverging from a point source located a distance  $z$  from the grating plane. The path length of a ray traveling from the source to a point  $x$  is  $R$ , as shown in Fig. 3.  $\theta$  and  $\phi$  are defined as the spherical polar and azimuthal angles, and the  $x$ -axis is defined perpendicular to the direction of the grating lines.

In the plane of the grating, the radial coordinate,

$$r = z \tan \theta, \quad (101)$$

and  $x$  in the new coordinate system is

$$x = r \cos \phi = z \tan \theta \cos \phi. \quad (102)$$

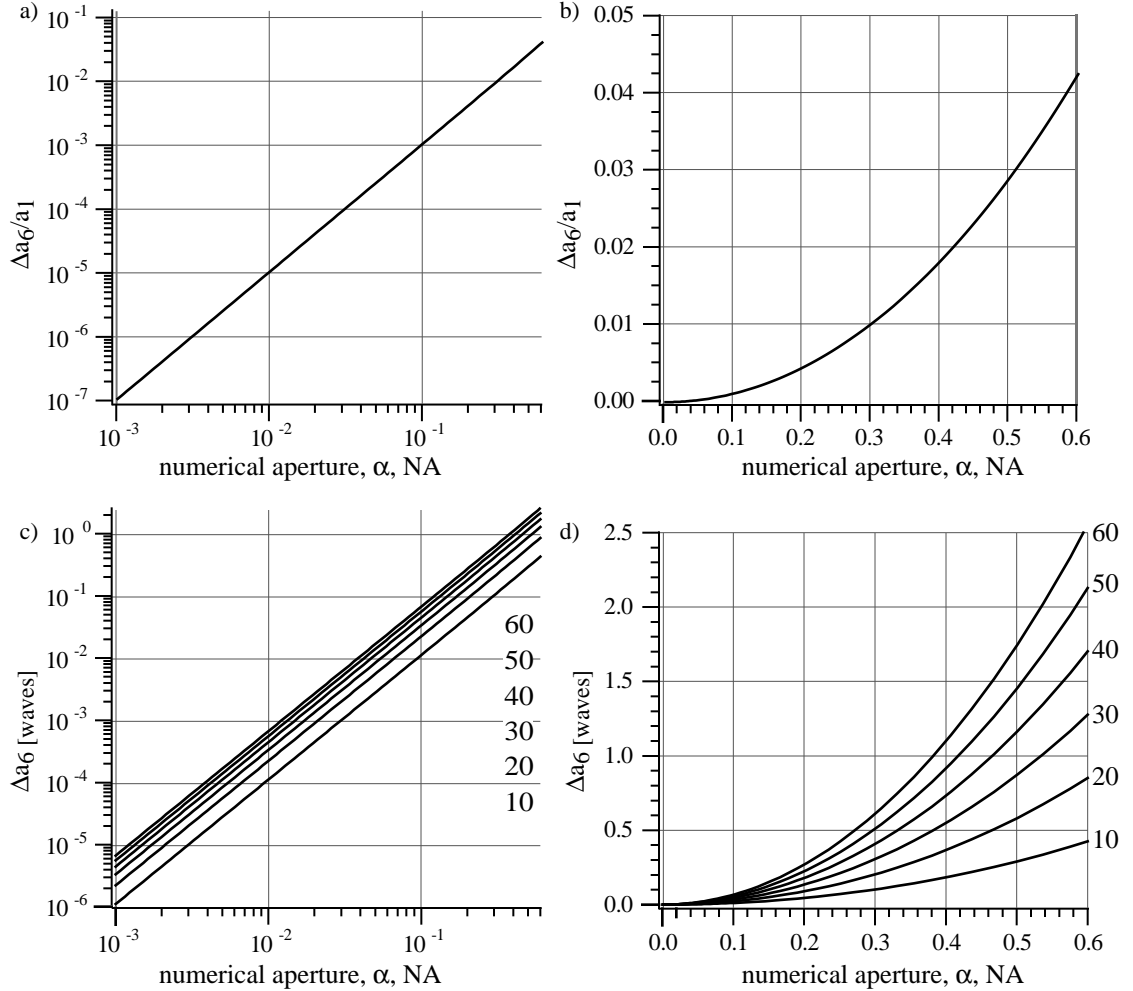
The path length  $R$  from the source point to the grating is

$$R = z + r \sin \theta = z + z \tan \theta \sin \theta = z(1 + \tan \theta \sin \theta). \quad (103)$$

Our immediate goal is to discover the phase of a beam that interferes with the illuminating beam to produce the grating pattern in intensity. As before, we assume the two interfering beams are of uniform intensity across the illuminated area. We may express the grating transmission function in the new coordinate system, using  $\kappa$  defined as before:  $\kappa \equiv 2\pi/d$ .

$$T(x) = \frac{1}{2} + \frac{1}{2} \cos(\kappa x) = \frac{1}{2} + \frac{1}{2} \cos(\kappa z \tan \theta \cos \phi). \quad (104)$$

The normalized intensity pattern produced by the interference of two beams is



**Figure 19.** The amount of systematic coma error introduced by the planar grating in a spherical beam depends on the system NA in the vicinity of the grating, and on the amount of tilt or equivalently the number of fringes in the measurement. The top row shows the ratio of the Zernike coma coefficient to the tilt coefficient plotted in log-log and linear scales. The lower row calculates the amount of coma for a certain number of interferogram fringes within the measurement NA. The same information is plotted in log-log and linear scale. The grating angle is arbitrary and therefore this discussion is easily extended to the  $a_7$  coma and the  $a_2$  tilt components.



$$I(\mathbf{r}) = \frac{1}{4} \left| e^{ikR} + e^{i\Phi} \right|^2 = \frac{1}{2} + \frac{1}{2} \cos(\Phi - kR). \quad (105)$$

Comparing Eqns. (104) and (105) allows us to solve for the phase function  $\Phi$ .

$$\Phi = kR + \kappa x = kR + \kappa z \tan \theta \cos \phi. \quad (106)$$

Here the  $kR$  term is not expanded in order to keep the form of the diverging (or converging) beam in the expression of interfering beam. The remaining term has only  $x$  dependence because of the  $\cos \phi$  component. A series expansion in  $\theta$  reveals separate terms that represent the spherical component, the diffraction angle of the second beam, and higher-order phase aberrations.

$$\Phi = kR + \kappa z \cos \phi \left[ \theta + \frac{1}{3} \theta^3 + \frac{2}{15} \theta^5 + \dots \right] = kR + (\kappa z \cos \phi) \theta + \frac{1}{3} (\kappa z \cos \phi) \theta^3 + \frac{2}{15} (\kappa z \cos \phi) \theta^5 + \dots \quad (107)$$

$$\Phi = (\text{spherical part}) + (\text{tilt}) + (\text{higher - order aberrations}). \quad (108)$$

As a final step, it is convenient to represent the phase terms in the *Normalized Beam Coordinate System* where the polar angle  $\theta$  is normalized to the NA angle  $\alpha$ :  $\rho \equiv \theta/\alpha$ . Here,  $\rho$  is a dimensionless *angular radius* variable, that allows us to make the transition from a *Beam Coordinate System*, to a representation on a unit circle, over which the Zernike Polynomials are orthogonal. Here it is important to remember that  $\alpha$  is the *local NA* angle describing the optical system in the vicinity of the grating. Clearly, if the beam is planar (collimated) as it reaches the grating, then  $\alpha = 0$ , and there is no systematic effect introduced by the grating, regardless of the image-side NA.

$$\Phi = kR + (\alpha \kappa z) \rho \cos \phi + \left( \frac{1}{3} \alpha^3 \kappa z \right) \rho^3 \cos \phi + \left( \frac{2}{15} \alpha^5 \kappa z \right) \rho^5 \cos \phi + \dots \quad (109)$$

We can write this explicitly in terms of the Zernike polynomials, as described in Chapter 14, using the shorthand notation for the Zernike polynomials  $Z_i \equiv Z_i(\rho, \phi)$ . After tilt, the third and fifth-order  $x$ -direction coma terms are

$$Z_1 = \rho \cos \phi, \quad (110a)$$

$$Z_6 = (3\rho^3 - 2\rho) \cos \phi, \quad (110b)$$

$$Z_{13} = (10\rho^5 - 12\rho^3 + 3\rho) \cos \phi. \quad (110c)$$

The isolated cubic, and fifth-order terms that appear in Eq. (109) can be re-written using

$$\rho^3 \cos \phi = \frac{1}{3} Z_6 + \frac{2}{3} Z_1, \quad (111a)$$

and 
$$\rho^5 \cos \phi = \frac{1}{10} Z_{13} + \frac{2}{5} Z_6 + \frac{1}{2} Z_1. \quad (111b)$$

Keeping only terms up to fifth-order, we can now rearrange terms to write Eq (109) as

$$\Phi = kR + \kappa z \alpha \left( 1 + \frac{2}{9} \alpha^2 + \frac{1}{15} \alpha^4 \right) Z_1 + \kappa z \alpha^3 \left( \frac{1}{9} + \frac{4}{75} \alpha^2 \right) Z_6 + \frac{2}{75} \kappa z \alpha^5 Z_{13} \quad (112)$$

Aside from the aberrations in the test optics, and other unrelated systematic error sources, this is the measured phase.

The methods for removing the grating coma follow directly from the removal of the systematic coma, described in Section 5.5. If  $\alpha$ , the *local NA* in the vicinity of the grating, is well-known then the removal may be straightforward. Note that it may be the case that the *measurement NA*, which includes only the sub-region of the beam involved in the analysis, is smaller than the available NA at the detector. In such cases, the  $\alpha$  used in the calculations should be the *measurement  $\alpha$* , representing only the subset of rays that eventually reach the detector and are used in the analysis.

At this point, the direction of the grating rulings can be generalized. The description is simplified by using the representation of Zernike pairs in vector notation, as described in Section 5.5.1. Here, we utilize a tilt vector  $\mathbf{T} \equiv (a_1, a_2)$ , a coma vector  $\mathbf{C} \equiv (a_6, a_7)$ , and we introduce a fifth-order coma vector  $\mathbf{C}_5 \equiv (a_{13}, a_{14})$ . By matching the coefficients in Eq. (112) the magnitude of the grating coma is proportional to the tilt

$$\Delta\mathbf{C} = \frac{\left(1 + \frac{12}{25}\alpha^2\right)\alpha^2}{9 + 2\alpha^2 + \frac{3}{5}\alpha^4} \mathbf{T} \approx \frac{1}{9}\alpha^2 \mathbf{T} . \quad (113)$$

Hence the adjustment to  $\mathbf{C}$  required to remove the grating coma is

$$\mathbf{C}' = \mathbf{C} - \Delta\mathbf{C} . \quad (114)$$

The fifth-order correction  $\mathbf{C}_5$  is always more than one order of magnitude smaller than the third-order correction.

$$\Delta\mathbf{C}_5 = \frac{\frac{2}{75}\alpha^4}{1 + \frac{2}{9}\alpha^2 + \frac{1}{15}\alpha^4} \mathbf{T} \approx \frac{2}{75}\alpha^4 \mathbf{T} , \quad (115)$$

and the required adjustment is  $\mathbf{C}'_5 = \mathbf{C}_5 - \Delta\mathbf{C}_5$  . (116)

Figures 19(a) and (b) shows the magnitude of  $\Delta a_6$  relative to  $a_1$ , calculated for between 10 and 60 fringes. The magnitude in waves is plotted in Figs. 19(c) and (d).

If the measurement NA is not well known, then a method of combining orthogonal measurements, as described in Section 5.5.2, should be employed.

### 5.9 Summary

**Grating Coma.**  $|\mathbf{C}| \approx 1/9 \text{ NA}^2 |\mathbf{T}| = 3.6 \times 10^{-4}$  waves per fringe.  $|\mathbf{C}| = 0.19 \text{ nm @ 40 fringes.}$

### 5.10 SPATIAL FILTERING BY THE IMAGE-PLANE WINDOW

By passing the test beam through a finite window in the image plane, the PS/PDI performs an inher-

ent spatial filtering of the light in a way that the PDI does not. The significance of this low-pass filtering depends on the size and shape of the window relative to the size of the focused beam. The relevant length scale, it will be shown, is  $\lambda/\text{NA}$ .

A certain amount of filtering is required to ensure that the overlap of the adjacent orders is minimized as the test beam passes through the window. If the system is designed carefully, then with the reference beam centered on the reference pinhole, the test beam passes through the center of the window.

Since the filter sits in the image plane of the test optic, and measurements are performed in the far-field, the window may be regarded simply as a spatial filter in the *Fourier domain* of the beam. This description is represented in Fig. 20. For the test beam, the window acts as a broad, low-pass filter. The pinhole acts as a *very narrow* low-pass filter for the reference beam (ideally, a delta-function). The window, displaced significantly from the central ray of the reference beam, functions as a band-pass filter, transmitting, or *leaking*, higher spatial-frequency components. The fact that these effects are readily observed in the data has led to the development of an alignment system based on a rapid 2-D Fourier-transform of the measured data. These observations are discussed in Section 6.5.)

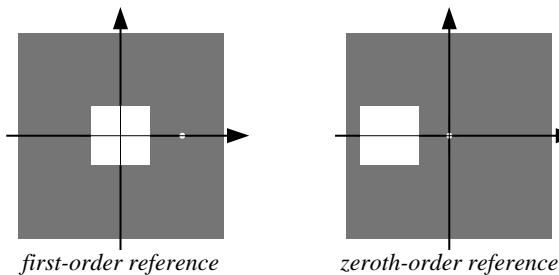
#### 5.10.1 A Simple Model for Spatial Filtering

This section presents a simple mathematical treatment of the window's spatial filtering effect. Based on the fact that the light propagates from the exit pupil of the test system to a focus in the image-plane, and then to the detector in the far-field, we may regard the pinhole and window as spatial filters in the Fourier domain of the beam, as stated above. For spatial filters of moderate dimension, and sufficient distance to the detector, the far-field (Fraunhofer) approximation for the diffraction calculations is suitable (Goodman 1988:61). The near-field term becomes significant only when for the lateral distance  $r$ ,

$$r \geq \sqrt{\frac{\lambda z}{\pi}} . \quad (117)$$

This is approximately  $20 \mu\text{m}$  at  $13.4\text{-nm}$  wavelength and  $10\text{-cm}$  distance.

Define  $G_i$  as the electric field of the test beam in the exit pupil of the test optic.  $G_d$  is the test beam field as it reaches the detector, after having passed through the image plane. Let  $g$  be the field in the image-plane, and  $t$  be the transmission function of the window; either or both may be complex. Here we



**Figure 20.** A simple lateral translation of the image-plane spatial filter in the PS/PDI switches between the first-order reference and the zeroth-order reference configurations. The axes here are centered on the test beam focus. The positive and negative first-diffracted-orders fall on opposite sides of the focus. Translating the spatial-filter makes either the zeroth-order beam or one of the first-order beams the reference beam.

will use the symbol  $F\{\}$  to denote the Fourier-transform in the following manner:

$$F\{b(\mathbf{k})\} \equiv \int_A b(\mathbf{k}) e^{i\mathbf{k} \cdot \mathbf{r}} d\mathbf{k} \equiv B(\mathbf{r}) . \quad (118)$$

Based on our assumptions for  $G_i$  and  $g$ ,

$$F\{G_i\} = g, \quad F\{T\} = t . \quad (119)$$

The Convolution Theorem allows us to determine  $G_d$

$$F^{-1}\{g \cdot t\} = G_i \circ T \equiv G_d . \quad (120)$$

The features of  $G_i$  are essentially mapped onto  $G_d$ . The effect of the filtering appears in the detected field as a convolution of the propagated test beam and the Fourier-transform of the window transmission function.

The effect of an arbitrary filter may be studied in terms of its transform  $T$ . When the window is small, then the central peak of  $T$  is angularly broad, and the convolution of  $G_i$  with  $T$  blurs any sharp features in  $G_i$ , decreasing the spatial frequency content of the measurement. Otherwise, when the window is large, the peak of  $T$  will be very narrow, and the convolution of  $G_i$  with  $T$  will leave  $G_i$  largely unaffected.  $T$  helps us to define the *angular resolution* of a given filter.

Let us consider a square window of width  $w$ , and explicitly write the diffraction equation. Some leading constant coefficients are ignored for simplicity.

$$T(\mathbf{k}) = \int_{w/2}^{w/2} t(\mathbf{r}) e^{i\mathbf{k} \cdot \mathbf{r}} d\mathbf{r} = \int_{w/2}^{w/2} e^{ik_x x} dx \int_{w/2}^{w/2} e^{ik_y y} dy . \quad (121)$$

$$T(\mathbf{k}) = w^2 \operatorname{sinc}\left(\frac{1}{2} w k_x\right) \operatorname{sinc}\left(\frac{1}{2} w k_y\right) . \quad (122)$$

$T$  may be expressed in terms of the polar angles in the  $x$ - and  $y$ -directions  $\boldsymbol{\theta} \equiv (\theta_x, \theta_y)$ .

$$T(\boldsymbol{\theta}) = w^2 \operatorname{sinc}\left(\frac{1}{2} w k \theta_x\right) \operatorname{sinc}\left(\frac{1}{2} w k \theta_y\right) . \quad (123)$$

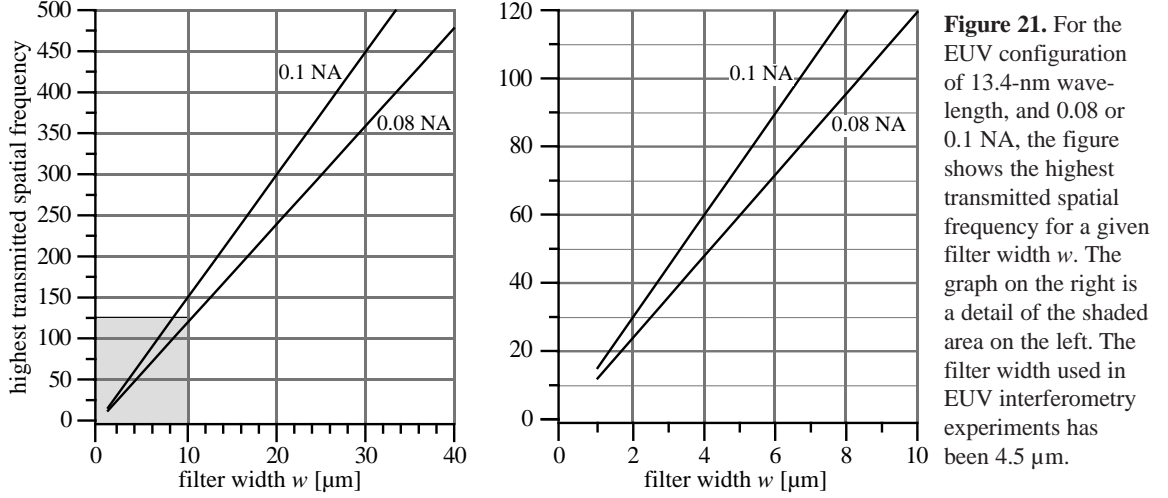
The full-width of the central peak of  $\operatorname{sinc}(x)$  is approximately  $\pi$ . Thus, the angular width of  $T$  is given approximately by

$$\frac{w k \Delta\theta}{2} = \pi \Rightarrow \Delta\theta = \frac{2\pi}{w k} = \frac{\lambda}{w} . \quad (124)$$

This width itself has no dependence on NA. However, its significance on the highest measurable spatial frequencies, is given by the ratio of the full angle of the optical system  $2\alpha$  to  $\Delta\theta$ .

$$\text{highest frequency} = \frac{2\alpha}{\Delta\theta} = \frac{2\alpha w}{\lambda} = \frac{2w}{\lambda / \text{NA}} \text{ cycles} . \quad (125)$$

Features of higher spatial frequency than this will subtend an angle smaller than  $\Delta\theta$ ; the convolution will strongly blur these features.  $\Delta\theta$  may be regarded as the *angular resolution* of the PS/PDI in any direction. For the EUV configuration of 13.4-nm wavelength, with 0.08 or 0.1 NA, Fig. 21 shows the highest trans-



mitted spatial frequency for a range of reasonable window sizes.

### 5.10.2 Effect of Spatial Filtering on the Intensity and Phase Measurement

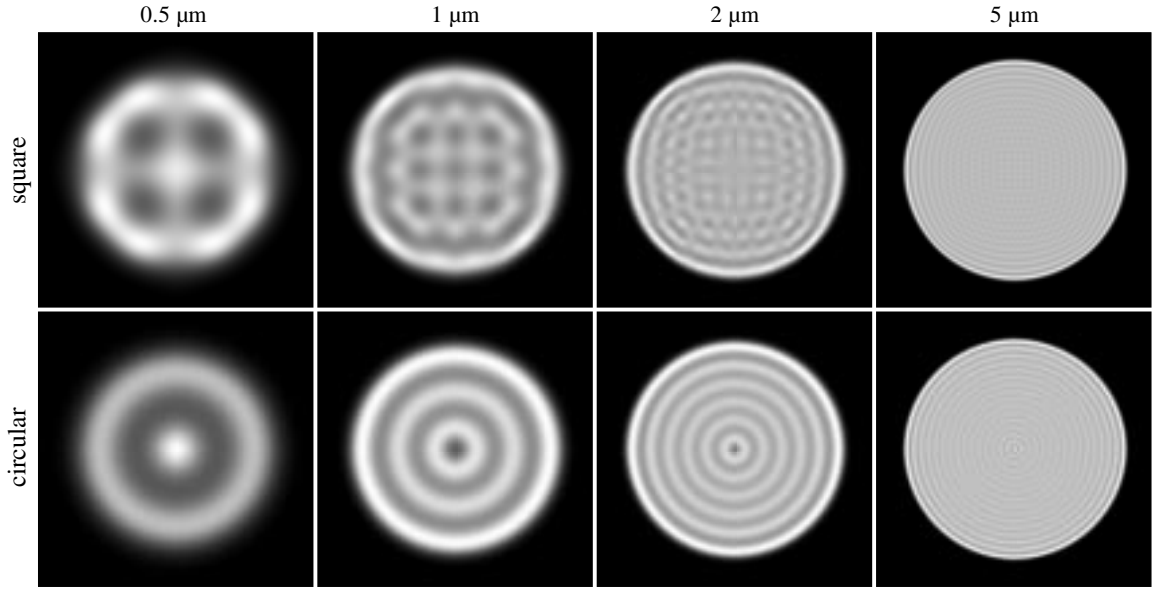
The above description showed that the image-plane spatial filtering of the window may *blur* sharp (angularly narrow) features in the test wavefront when the window is relatively small. It is important to understand how this blurring may affect the measured intensity *and phase* of the test wavefront. For every specific design of the window's size, shape, and position, this effect will be somewhat different. In this section, an informal, heuristic argument provides a useful tool for demonstrating that when the test optics are of high quality, intensity ripples adjacent to sharp features may not be accompanied by ripples in the phase.

If the centered window transmission function  $t(\mathbf{r})$  is strictly real and has polar symmetry, then its Fourier-transform  $T(\mathbf{k})$  is also strictly real. The following equation demonstrates this point for an arbitrary real function  $t(\mathbf{r})$ .

$$T(\mathbf{k}) = \int_A t(\mathbf{r}) e^{i\mathbf{k}\cdot\mathbf{r}} d\mathbf{r} = \int_{\frac{A}{2}} t(\mathbf{r}) (e^{i\mathbf{k}\cdot\mathbf{r}} + e^{-i\mathbf{k}\cdot\mathbf{r}}) d\mathbf{r} = 2 \int_{\frac{A}{2}} t(\mathbf{r}) \cos(\mathbf{k}\cdot\mathbf{r}) d\mathbf{r} \in \mathbf{R}. \quad (126)$$

Depending on the shape of the window,  $T$  may have a series of positive and negative lobes. (This is the case for the rectangular window and its accompanying sinc function transform. Furthermore, in an aberration-free optical system, apart from any constant coefficient the test beam  $G_i$  is also real. Thus the convolution of the test beam  $G_i$  with  $T$  (that is,  $G_d$ ) is real.

Sharp changes in the test beam intensity may occur where there are physical apertures or pupils within the system, or where defects in the optical surfaces create localized dark regions. At the detector, these sharp changes in the intensity may be accompanied by intensity oscillations, due (mathematically) to the convolution of the test beam with the lobes of  $T$ . As  $T$  is a strictly real function, however, there is *no variation* in the phase of  $G_d$ . Clearly, if the test optic contains features that create rapid *phase* variations across the aperture then there will be accompanying ripples in phase as well.



**Figure 22.** Simulated image-plane spatial filtering of a uniformly illuminated optical system with a circular aperture. The window in the PS/PDI spatial filter transmits the test beam with some spatial filtering. For 13.4 nm wavelength and 0.08 NA, the effect of variously sized square and circular filters is shown. The intensity recorded in the detector plane clearly shows the effects of filtering at 5- $\mu\text{m}$ -width and below..

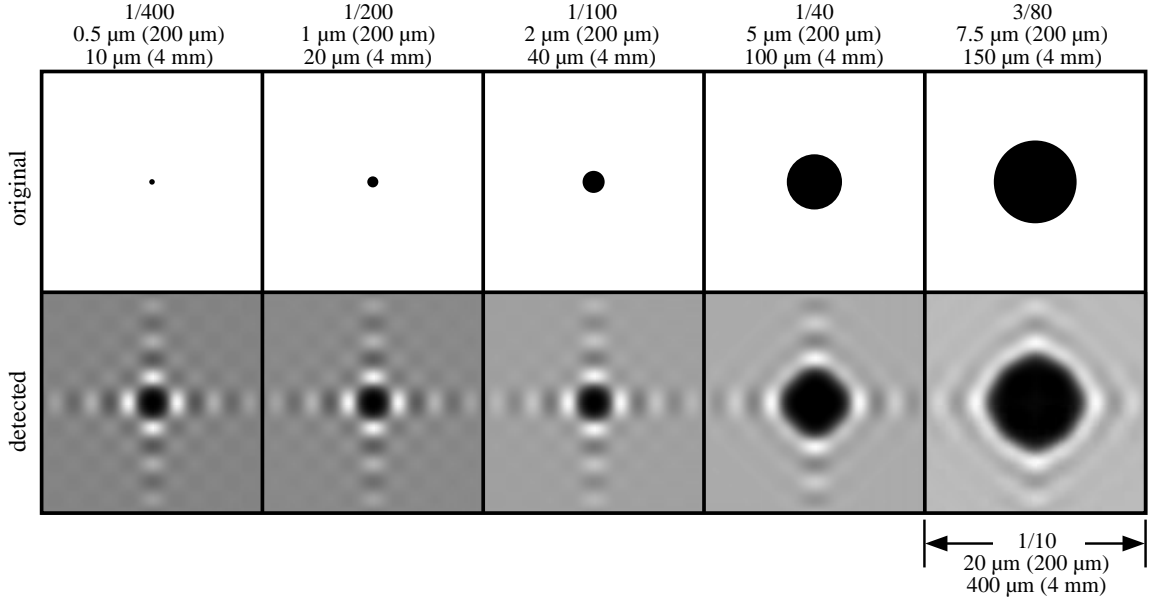
### 5.10.3 Examples

A straightforward diffraction simulation is performed to illustrate the effect of various amounts of spatial filtering on a PS/PDI interferometer for EUV optical system measurement. Considering an aberration-free optical system operating at 13.4 nm wavelength with 0.08 NA, both circular and square windows of different widths are studied.

Figure 22 shows the (simulated) detected intensities for square (top row) and circular (bottom row) windows 0.5- to 5.0- $\mu\text{m}$ -wide. As described above, the ripples are caused by the convolution of the circular pupil with the Fourier-transform of the window transmission function.

Another subject of interest is the way in which spatial filtering affects small, localized defects in the optic. Figure 23 shows the results of a simulation in which these defects are modeled as dark circles in an otherwise bright region far from the edge of the aperture. The top row shows how these sharply-defined dark regions appear in the exit pupil. At 0.08 NA, the fraction of the whole aperture subtended by these features is shown above each. For reference, these relative sizes are also provided for the two cases of the zoneplate and the Schwarzschild objective experiments. The pupil sizes in these cases are 200  $\mu\text{m}$  and 4 mm respectively. The relevant dimension of the dark features is their angular size with respect to the full aperture of the test optic. The bottom row shows the (simulated) intensity patterns at the detector, scaled for display. The *angular width* of these simulation images is 1/10-th of the aperture.

Notice that below 1/40-th of the aperture width, the features are below the angular resolution of the



**Figure 23..** Spatial filtering of defects in the test optic is simulated in this figure. Here, defects are modeled as dark solid circles in an otherwise bright pupil illumination pattern. Above each simulation, the defect sizes are given as a fraction of the full angular width of the optic, with relevant numbers quoted for the zoneplate experiment and the 10× Schwarzschild objective. Here the defect size is given with the full aperture size shown in parentheses. The filter width is 5 μm. Details of the same angular area of the pupil as recorded in the detector plane are shown in the lower row; the images are scaled for display. Below 1/40-th of the full-angle, the defects behave essentially as delta-function aberrations, unresolved by the spatial filter. Above 1/40-th of the full-angle, the recorded test beam patterns follow the increasing angular size of the defects. The horizontal and vertical pattern in the test beam images is the sinc function generated by the square window shape.

window, and all appear very similar. Because of their relatively small size, the resultant intensity patterns reveal the behavior of  $T$ . Mathematically this situation is analogous to the convolution of a delta-function with  $T$ . In qualitative agreement with Eq. (125) plotted in Fig. 21, the highest transmitted frequency lies between 1/40 and 1/100 of the aperture width. Features smaller than this size are not *resolved* by the filter.

### 5.10 Summary

**Spatial Filtering.** Highest spatial frequency  $f = 2w/(\lambda/NA)$  cycles  $\Rightarrow$  12 cycles/μm filter width.

### 5.11 VARIATIONS OF THE PS/PDI SPATIAL FILTER

The previous section described the way in which the size of the image-plane window affects the highest spatial frequencies resolvable with a given configuration of the PS/PDI. Along the window-pinhole displacement direction, the maximum allowable width of the window is constrained by the image-plane separation of the test and reference beams. However, in the perpendicular direction, there is no constraint on the size of the window: the window may be defined as a long slit, narrow in the displacement direction.

The square window design is easily generalized to the case of a rectangular window of dimensions  $w_x$  and  $w_y$ . The Fourier-transform of the rectangular window transmission function is

$$T(\boldsymbol{\theta}) = w_x w_y \text{sinc}\left(\frac{1}{2} w_x k \theta_x\right) \text{sinc}\left(\frac{1}{2} w_y k \theta_y\right). \quad (127)$$

Separating the  $x$ - and  $y$ - directions allows the definition of two angular convolution half-widths.

$$\Delta\theta_x = \frac{\lambda}{w_x}, \text{ and } \Delta\theta_y = \frac{\lambda}{w_y}. \quad (128)$$

These serve as the  $x$ - and  $y$ -direction angular resolution of the system. As described in Section 5.10, to maintain a high spatial frequency response, it is desirable to keep  $\Delta\theta$  as small as possible. It is not necessary, however, to reduce  $\Delta\theta$  significantly beyond the angular resolution of the detector, usually determined by the pixel size and the detector placement.

To minimize overlap of adjacent orders on either side of the test beam, the size of the window is constrained in the displacement direction. The width should not exceed the beam separation distance  $s$ . With  $\mathbf{x}$  aligned parallel to the displacement, this constraint on the maximum size of  $w_x$  ( $w_x < s$ ) limits the minimum achievable  $\Delta\theta_x$ . In the  $y$ -direction, since there is no such constraint,  $\Delta\theta_y$  may be made as small as desired.

#### 5.11.1 Image-Plane Window/Pinhole Filter Designs

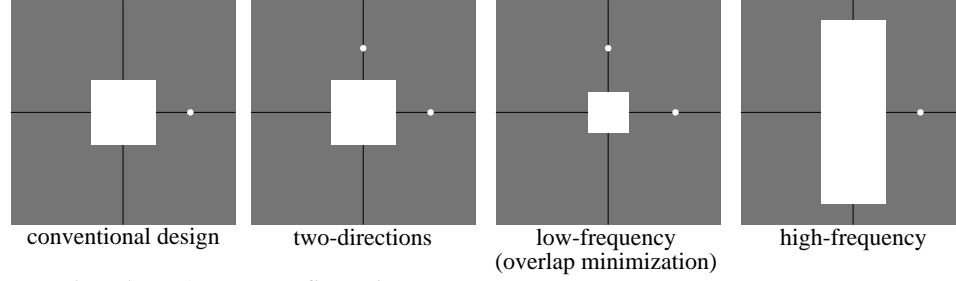
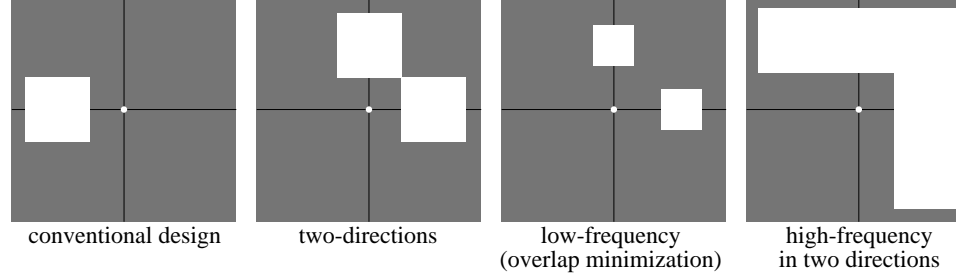
There are a number of available designs for the window and pinhole spatial filters some of which are shown in Fig. 24. These designs, each allows only two beams to pass through at a time, are divided into two sets to distinguish between the *first-order reference* and the *zeroth-order reference* configurations. By definition, the reference beam is whichever beam is filtered by passing through the small reference pinhole.

Several designs enable measurement in two directions without requiring window translation. In two measurements, one grating may be replaced by another, oriented with its rulings rotated by 90 degrees from the first. Sections 5.5 describes the importance of having separate measurements performed with different test and reference beam displacements.

One advantage of the first-order reference configuration not previously addressed is the fixed position of the test beam when separate measurements are preformed. This guarantees that the same field point is being measured. Plus, as described in Sections 5.8 and 5.9, the first-order reference configuration can be used to filter aberrations introduced by a grating beamsplitter. The first-order reference *two-directions* design shown in Fig. 24 was chosen for the experiments in EUV interferometry described in this thesis. The ability to perform a pair of measurements without translating the beam is very important if the beam separation distance is significant with respect to the field-of-view of the test optic.

In the presence of large mid- or high-spatial frequency aberrations, which scatter light away from the central peak in the image-plane, it may be desirable to reduce the amount of beam overlap, by

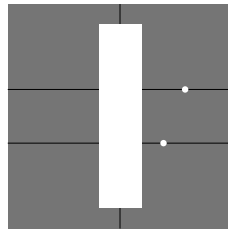


First-order reference configurationsZeroth-order reference configurations

**Figure 24.** Several designs for the PS/PDI image-plane spatial-filter in both the first-order reference and the zeroth-order reference configurations. Patterns that are symmetric about  $\theta = 45^\circ$  are designed to be used with two separate  $90^\circ$  orientations of the beamsplitter. Larger windows transmit a greater range of test beam spatial frequencies, however *leakage* of the reference beam through the windows may introduce measurement uncertainties.

decreasing the size of the window. The so-called *low-frequency* designs shown in Fig. 24 sacrifice spatial resolution to improve data quality.

As described above, there is no constraint on the size of the window perpendicular to the beam separation direction. The *high-frequency* designs shown in Fig. 24 exploit this fact by using a rectangular window, long in one direction, to transmit high-spatial-frequencies. In the first-order reference configuration, however, it may not be possible to have two orientations of measurement with a single high-frequency design. The filter design shown in Fig. 25 achieves the objectives of having two different beam separations *and* high-spatial-frequency response (in one direction) with either the zeroth-order reference or the first-order reference configurations. Two gratings of different pitch but oriented in the same direction may be placed on the same translation stage to simplify the experimental apparatus.



**Figure 25.** An image-plane spatial filter design that allows measurement with two different beam separations, and provides high-spatial-frequency response in both the zeroth-order and the first-order reference configurations. Two gratings of different pitch, but same orientation are used. The reference pinholes are intentionally displaced to avoid overlap from adjacent diffracted orders.

### 5.11 Summary

- **Filter Design.** Place pinholes at  $90^\circ$  adjacent to a square window to enable direct measurements of systematic effects. Separately, adjust width perpendicular to beam separation to improve spatial frequency response.

### 5.12 DISTORTIONS DUE TO THE PLANAR DETECTOR

In the absence of re-imaging optics, the test and reference beams propagate as spherically diverging beams incident on a planar detector. Previous sections (5.5 and 5.6) have described the systematic error contributions of the beam separation at focus, and detector misalignment, based on the geometry of the system. Another source of systematic error is the small geometric distortion arising from the planar geometry of the detector itself, intercepting the spherical beams.

Unlike the previous systematic error components which arise from a path-length difference between the test and reference beams, this effect may be described as a systematic, radial distortion across the measured area. In the angular representation of the *Beam Coordinates*, the planar *Detector Coordinate System* becomes non-linear with a purely radial dependence.

In the *Beam Coordinate System*, the polar angle at a given detector position is  $\theta$ .  $r$  is the radial detector position.

$$r(\theta) = z \tan \theta \Leftrightarrow \theta(r) = \tan^{-1}\left(\frac{r}{z}\right). \quad (129)$$

The radius  $r_\alpha$  in the detector plane corresponds to rays at the maximum angle within the NA of measurement  $\alpha$ ,

$$r_\alpha = z \tan \alpha. \quad (130)$$

As before,  $t$  is the tangent of the NA angle  $\alpha$ .

$$\frac{r_\alpha}{z} = \tan \alpha \equiv t. \quad (131)$$

In the *Normalized Detector Coordinates*, the dimensionless radius  $\rho$  is defined as

$$\rho \equiv \frac{r}{r_\alpha}. \quad (132)$$

Now,  $\theta$  may be rewritten in terms of these new parameters

$$\theta(r) = \tan^{-1}\left(\frac{r}{z}\right) = \tan^{-1}\left(\frac{r_\alpha}{z} \frac{r}{r_\alpha}\right) = \tan^{-1}(tp) = \theta(\rho). \quad (133)$$

When a measurement is made, the wavefront is typically sampled on an array linear in  $r$  (or  $\rho$ ). Equation (133) represents a *correction* which must be performed after measurement, to put the wavefront back into its natural, spherical coordinate system. To make the transition from the normalized *Detector* coordinate system to a normalized *Beam* coordinate system, we divide  $\theta$  by  $\alpha$  as described in Section 5.2. Define  $\gamma$  as the normalized polar angle.

$$\gamma(\rho) = \frac{\theta(\rho)}{\alpha} = \frac{1}{\alpha} \tan^{-1}(tp). \quad (135)$$

By their definitions,  $\gamma$  and  $\rho$  will be equal only at the central point ( $\gamma = \rho = 0$ ) and at the edge of the mea-

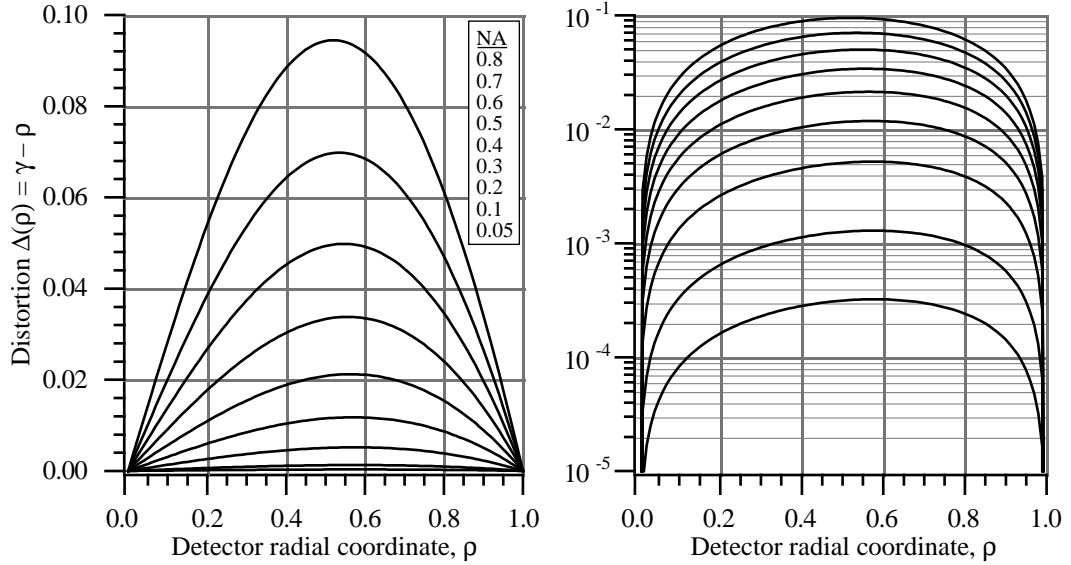
surement NA ( $\gamma = \rho = 1$ ). For the domain of  $\rho$  on  $(0,1)$ , there is some distortion, dependent on the NA. For small measurement NA,  $\gamma$  approximates  $\rho$  over the entire range. For large measurement NA, however, the non-linearity causes the two to diverge. This dimensionless distortion may be characterized by introducing a parameter  $\Delta$  defined as

$$\Delta(\rho) = \gamma(\rho) - \rho = \frac{1}{\alpha} \tan^{-1}(\rho \alpha) - \rho = \frac{1}{\alpha} \tan^{-1}(\tan \alpha \rho) - \rho. \quad (136)$$

$\Delta$ , which is defined in the normalized coordinate system, indicates the difference between the actual polar angle (normalized) and the radial position on the detector. In one interpretation, for a given  $\rho$ ,  $\Delta$  represents the amount of *radial shift* that is required to remove the distortion. Since by definition  $\gamma$  and  $\rho$  agree at 0 and 1,  $\Delta$  must be zero at these points. Figure 26 shows  $\Delta(\rho)$  plotted for nine different values of NA (recall,  $\alpha \equiv \text{NA}$ ). Table 3 first lists the peak value of the distortion for each NA shown in Fig. 26, and then translates that number into pixels in several experimental measurement domains. With  $N$  as the pixel-width of a measured interferogram, the normalized peak distortion  $\Delta_{\text{peak}}$  is multiplied by the radius  $N/2$  to calculate the amount of distortion in pixels. Note that at 0.08 NA, the approximate image-side NA of present EUV lithographic optical systems, the peak distortion is  $8.22 \times 10^{-4}$ , less than 0.1%. This indicates the presence of a tiny, 0.33-pixel peak distortion in a typical, 800-pixel-diameter measurement.

### 5.12 Summary

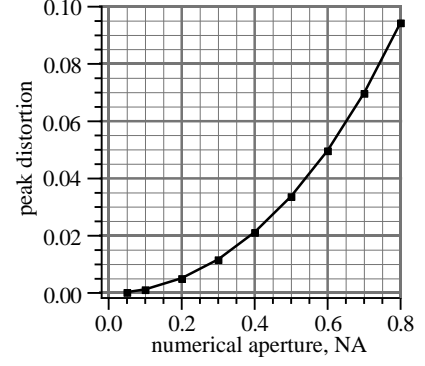
**Planar Detector Distortion.** For 800-pixel measurement diameter, and 0.08 NA, peak distortion  $\Delta = 8.22 \times 10^{-4} \Rightarrow 0.33$  pixels.



**Figure 26.** The use of a planar detector to record the spherically diverging interference pattern introduces a radial distortion into the coordinate system of measurement. As the radial position of a point on the detector is translated into an angular position in the beam, the two coordinate systems match only at the center and the domain edge. Within the array, the radial distortion  $\Delta(\rho)$  is defined in Eq. (136).

**Table 3.** Peak measurement distortion (in *pixels*) for a various array sizes, at different NA.  $\Delta_{peak}[\text{pixels}] = \Delta_{peak} * (N_{pixels}/2)$ .

NA	Distortion	Measurement Array Size (pixels)			
		250	500	750	1000
0.05	$3.20 \times 10^{-4}$	0.04	0.08	0.12	0.16
0.1	$1.29 \times 10^{-3}$	0.16	0.32	0.48	0.65
0.2	$5.17 \times 10^{-3}$	0.65	1.29	1.94	2.59
0.3	$1.18 \times 10^{-2}$	1.48	2.95	4.43	5.9
0.4	$2.12 \times 10^{-2}$	2.65	5.3	7.95	10.6
0.5	$3.38 \times 10^{-2}$	4.23	8.45	12.68	16.9
0.6	$4.98 \times 10^{-2}$	6.23	12.45	18.68	24.9
0.7	$6.98 \times 10^{-2}$	8.73	17.45	26.18	34.9
0.8	$9.45 \times 10^{-2}$	11.81	23.63	35.44	47.25



**Figure 27.** Peak distortion as a function of NA. The distortion is based on a unit-circle coordinate system.

### 5.13 SUMMARY OF SYSTEMATIC ERRORS AND RECOMMENDATIONS

The following list enumerates the most important results and systematic error effects described in this chapter. The numbers pertain to the at-wavelength measurement of an EUV lithographic optic operating at 13.4-nm wavelength with 0.08 NA.

- **5.3 Beam Separation.**  $s/N_{fringe} = \lambda/2t = 0.084 \mu\text{m}/\text{fringe} \Rightarrow N_{fringe}/s = 2t/\lambda = 12 \text{ fringes}/\mu\text{m}$ .
- **5.4 Bandwidth.**  $W_g = 2.22 \times 10^{-7}$  @ 0.1% BW (Gaussian distribution). Fringe amplitude is reduced by  $2.22 \times 10^{-7}$  per wave<sup>2</sup> of aberration at this bandwidth.
- **5.5 Measured Geometric Coma.**  $|C| \approx 1/6 \text{ NA}^2 |T| \Rightarrow$  At 0.08 NA,  $|C|/|T| = 1/6 \cdot 0.08^2 = 0.0011$  waves per wave of tilt =  $5.5 \times 10^{-4}$  waves per fringe.  $|C| = 0.37 \text{ nm}$  @ 50 fringes. At 0.1 NA,  $|C|/|T| = 1/6 \cdot 0.1^2 = 0.0017$  waves per wave of tilt =  $8.3 \times 10^{-4}$  waves per fringe.  $|C| = 0.56 \text{ nm}$  @ 50 fringes.
- **5.6 Detector Misalignment.** P-V astigmatism  $A = s\gamma\text{NA}^2 \Rightarrow \sim 0.47 \text{ nm}/^\circ$  tilt. Also,  $A/\gamma N_{fringe} = \lambda \text{ NA}/2 = 0.54 \text{ nm}/^\circ$  tilt/fringe. The measured Zernike coefficient of astigmatism is half of this, or  $0.27 \text{ nm}/^\circ$  tilt/fringe.
- **5.7 Shear Angle.**  $\theta = \lambda/d =$  shear angle.  $d =$  grating pitch.  $\beta =$  half-angle over which reference wavefront is of arbitrarily high quality.  $\alpha = \text{NA}$ . Minimum requirement:  $\beta > \theta + \alpha$ .
- **5.8 Grating Fabrication Errors.** Recommendation: use the *first-order reference* configuration whenever grating fabrication error magnitudes are unknown, or are known to be comparable with the desired accuracy.
- **5.9 Grating Coma.**  $|C| \approx 1/9 \text{ NA}^2 |T| = 3.6 \times 10^{-4}$  waves per fringe.  $|C| = 0.19 \text{ nm}$  @ 40 fringes.
- **5.10 Spatial Filtering.** Highest spatial frequency  $f = 2w/(\lambda/\text{NA})$  cycles  $\Rightarrow 12 \text{ cycles}/\mu\text{m}$  filter width.
- **5.11 Filter Design.** Place pinholes at  $90^\circ$  adjacent to a square window to enable direct measurements of systematic effects. Separately, adjust width perpendicular to beam separation to improve spatial frequency response.
- **5.12 Planar Detector Distortion.** For 800-pixel measurement diameter, and 0.08 NA, peak distortion  $\Delta = 8.22 \times 10^{-4} \Rightarrow 0.33 \text{ pixels}$ .